

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 11:58:06 ; Search time 32.62 Seconds
(without alignments)
3474.464 Million cell updates/sec

Title: US-09-099-898-1
Perfect score: 453
Sequence: 1 GTGTGCGGATTGGTTAGC.....GGGAGCTGGAATAAACCT 453

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_GenSeq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	453	100.0	453	1	V84059
2	420.6	92.8	553	1	X51720
3	194	42.8	229	1	X41471
4	158.4	35.0	572	1	V84057
5	152.4	33.6	560	1	V84056
6	92.8	20.5	603	1	V84058
7	78	17.2	554	1	V84055
8	64.4	14.2	574	1	V84054
9	54.4	12.0	527	1	V84053
10	40.6	9.0	985	1	V44439
11	40.6	9.0	985	1	V64548
12	39	8.6	8438	1	Q73500
13	38.6	8.5	114955	1	X53491
14	36.6	8.1	3415	1	T36481
15	36.6	8.1	11820	1	V18130
16	36.2	8.0	2932	1	Q25388
17	35.2	7.8	329	1	V44425
18	35.2	7.8	329	1	V84534
19	34.8	7.7	30001	1	T61016
20	34.8	7.7	117213	1	V62176
21	34.8	7.7	30001	1	X05110
22	34.8	7.7	114955	1	X53491
23	34.4	7.6	201	1	N70195
24	34.2	7.5	1035	1	T49316
25	34.2	7.5	1035	1	V36430
26	34.2	7.5	1035	1	V39975
27	34.2	7.5	1035	1	V54609
28	34.2	7.5	1035	1	V39991
29	34	7.5	1687	1	T39798
30	34	7.5	384	1	V41906
31	34	7.5	1545	1	V19063
32	34	7.5	201	1	X19063
33	33.8	7.5	6854	1	T76903
34	33.8	7.5	535	1	V44428
35	33.8	7.5	535	1	V64537
36	33.6	7.4	24379	1	T93095
37	33.6	7.4	24379	1	V25925
38	33.6	7.4	1439	1	V44424
39	33.6	7.4	1439	1	V64533
40	33.6	7.4	1236	1	V72041
41	33.6	7.4	1929	1	V72046
42	33.6	7.4	1926	1	V72054
43	33.6	7.4	2019	1	V72059

C 44 33.6 7.4 2025 1 V72064 Fusion protein PNG
C 45 33.6 7.4 1998 1 V72075 Fusion protein (80

ALIGNMENTS

RESULT 1

ID V84059 standard; cDNA; 453 BP.
AC V84059;
DT 09-MAR-1999 (first entry)
DE cDNA encoding a cysteine rich soluble protein designated C23.
KW Cysteine rich soluble protein; CRSP; C23; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy; ss.
OS Homo sapiens.
FH Key
FT CDS
FT Location/Qualifiers
FT 47..373
FT /*tag= a
FT /product= C23
FT 47..100
FT /*tag= b
FT mat_peptide 101..370
FT /*tag= c
PN WO9858061-A1.
PD 23-DEC-1998.
PF 18-JUN-1998; U12236.
PR 09-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PR (SCHE) SCHERING CORP.
PR Franz-Bacon K, Gorman DM, McClanahan TK;
PR WPI: 99-095339/08.
PR P-PSDB; W87710.
CC New cysteine-rich soluble proteins - used to modulate proliferation,
CC differentiation, trafficking and development of cells, e.g. for
CC treating inflammation, cancer and degeneration
CC Claim 16; Page 18-19; 119pp; English.
CC The present sequence encodes a cysteine rich soluble protein (CRSP)
CC designated C23. CRSP proteins, and their (ant)agonists, are used to
CC modulate physiology, differentiation, trafficking and development
CC of cells (including those in culture), particularly cells of the
CC mammalian immune system. They are used for treatment of abnormal
CC proliferation (cancer, inflammation or degeneration), regeneration,
CC degeneration and atrophy. The proteins are also used to raise, or
CC detect, antibodies, to design oligonucleotides for library screening,
CC in drug screens and to isolate cognate receptors. The antibodies are
CC used for affinity purification of CRSP, to screen expression libraries,
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC to produce anti-idiotypic antibodies (useful for diagnosis), in
CC competitive drug screens, and as therapeutic modulators.
CC Sequence 453 BP; 74 A; 134 C; 158 G; 87 T;
SQ

Query Match 100.0%; Score 453; DB 1; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.7e-106;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGCGGATTGGTTAGCTAGCCACCGAGAGCGCTGCAGGATGAAGCTCTCTG 60
DB 1 GTGTGCGGATTGGTTAGCTAGCCACCGAGAGCGCTGCAGGATGAAGCTCTCTG 60
QY 61 TCTG 120
DB 61 TCTG 120
QY 121 AGAAGCCATCATGATGAGAGGATCCAGAGGTCCCGGCTCCCTAATATTATGGGCAATAAG 180
DB 121 AGAAGCCATCATGATGAGAGGATCCAGAGGTCCCGGCTCCCTAATATTATGGGCAATAAG 180
QY 181 CAGCATTGCGCTGGAGTGCACGAGCGTCCACTCCAGGGGGACCTGCTACTTGCCTCCG 240

Db 181 CAGCATTTGGCTGGAGTGCAGAGCGCTACCTCCAGGGGGAGCTGGCTACTTGGCCCCCG 240
 QY 241 AGGCTTCGGCGTACCGGCTGCACTTGTGGCTCCGGCTGTGGCTCGTGGGATGTGGCGC 300
 Db 241 AGGCTTCGGCGTACCGGCTGCACTTGTGGCTCCGGCTGTGGCTCGTGGGATGTGGCGC 300
 QY 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 360
 Db 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 360
 QY 361 TGTGACGCTTGAAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 420
 Db 361 TGTGACGCTTGAAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 420
 QY 421 CGAGGGGTTGGCGGGAGCTGGAATAAACCT 453
 Db 421 CGAGGGGTTGGCGGGAGCTGGAATAAACCT 453

RESULT 2

ID X51720 standard; DNA; 553 BP.
 AC X51720:
 DT 17-JUN-1999 (first entry)
 DE DNA encoding a human secreted protein.
 KW Human secreted protein; cancer; immune disorder; infection;
 KW inflammatory disorder; skin disorder; tumour; atherosclerosis;
 KW restenosis; autoimmune disorder; Alzheimer's disease;
 KW peripheral neuropathy; trauma; spinal cord injury; allergy;
 KW hematopoietic disorder; skeletal disorder; neurological disorder;
 KW arthritic disorder; asthma; immunodeficiency disease; AIDS;
 KW transplant rejection; ss.
 OS Homo sapiens.
 PN W0911293-A1.
 PD 11-MAR-1999.
 PE 03-SEP-1998; U18360.
 PF 12-SEP-1997; US-058974.
 PR 05-SEP-1997; US-057626.
 PR 05-SEP-1997; US-057663.
 PR 05-SEP-1997; US-057669.
 PR 12-SEP-1997; US-058666.
 PR 12-SEP-1997; US-058667.
 PR 12-SEP-1997; US-058973.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Brexer-IA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
 PI Rosen GA, Ruben SM, Shi Y;
 DI WPI: 99-204988/17.
 DR P-PSDB; Y12933.
 PT New isolated human genes and the secreted polypeptides they encode
 PT - useful for diagnosis and treatment of e.g. neurological disorders,
 PT tumours, immune disorders, inflammation or haematological disorders,
 PS Claim 1; Page 165; 215pp; English.
 CC X51701-55 encode human secreted proteins. The polynucleotides and
 CC their corresponding secreted polypeptides are useful for preventing,
 CC treating or ameliorating medical conditions, e.g. by protein or gene
 CC therapy. Pathological conditions can also be diagnosed by determining
 CC the amount of the new polypeptides in a sample or by determining the
 CC presence of mutations in the new polynucleotides. Specific uses are
 CC described for each polynucleotide, based on which tissues they are
 CC most highly expressed in, and include developing products for the
 CC diagnosis or treatment of cancer, immune disorders, infection,
 CC inflammatory disorders, skin disorders, tumours, atherosclerosis,
 CC restenosis, autoimmune disorders, Alzheimer's disease, peripheral
 CC neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
 CC disorders, skeletal disorders, neurological disorders, arthritic
 CC disorders, asthma, immunodeficiency diseases, AIDS and transplant
 CC rejection. The polypeptides are also useful for identifying their
 CC binding partners.
 SQ Sequence 553 BP; 160 A; 132 C; 167 G; 94 T;

Query Match 92.8%; Score 420.6; DB 1; Length 553;
 Best Local Similarity 98.5%; Pred. No. 3e-98;

Matches 446; Conservative 0; Mismatches 4; Indels 3; Gaps 2;
 QY 1 GTGTGCGGATTTGGTTAGCTGAGCCACGAGAGGCGCTGCAGGATGAAGCTCTCTG 60
 Db 1 GTGTGCGGATTTGGTTAGCTGAGCCACGAGAGGCGCTGCAGGATGAAGCTCTCTG 60
 QY 61 TCT 120
 Db 61 TCT 120
 QY 121 AGAAGCCATCAATGAGAGGATCCAGGAGTCCGGGCTCCCTATATATTTAGGGCAATAAG 180
 Db 121 AGAAGCCATCAATGAGAGGATCCAGGAGTCCGGGCTCCCTATATATTTAGGGCAATAAG 180
 QY 181 CAGCATTTGGCTGAGTGCAGAGCGTCACTCCAGGGGAGCTGCTACTTCCGCCCG 240
 Db 181 CAGCATTTGGCTGAGTGCAGAGCGTCACTCCAGGGGAGCTGCTACTTCCGCCCG 240
 QY 241 AGGCTTCGCCGCTCACCGGCTGCACCTTGTGGCTCCGGCTGTGGGATGTGGCGGC 300
 Db 241 AGGCTTCGCCGCTCACCGGCTGCACCTTGTGGCTCCGGCTGTGGGATGTGGCGGC 300
 QY 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 360
 Db 301 CGAGACCATGTCACTGCGAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 360
 QY 361 TGTGACGCTTGAAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 420
 Db 361 TGTGACGCTTGAAGTGCAGGCGCATGGACTGGACCGGAGCGGCTGTGTGCG 420
 QY 421 CGAGGGGTTGGCGGGAGCTGGAATAAACCT 453
 Db 419 CGGA-GGGTTCGGGGAGCTGGAATAAACCT 450

RESULT 3
 X41471
 ID X41471 standard; cDNA; 229 BP.
 AC X41471;
 DT 22-JUN-1999 (first entry)
 DE Human secreted protein 5' EST SEQ ID NO: 130 from WO 9006553.
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; antiinflammatory; tumour inhibition; antitumour; ds.
 OS Homo sapiens.
 PN W0906553-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998; J1237.
 PR 01-AUG-1997; J1237.
 PA (GEST) GENSET.
 DR WPI: 99-153783/13.
 DR P-PSDB; Y12613.
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries derived from umbilical cord, lymph ganglia,
 PT lymphocytes and placental tissue
 PS Claim 1; Page 238; 41pp; English.
 CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y12521 to Y12668,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haematopoiesis regulating activity, tissue growth regulating
 CC activity, reproductive hormone regulating activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, antiinflammatory activity, tumour inhibition activity
 CC or other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for

Query Match 42.8%; Score 194; DB 1; Length 229;

Query Match 35.0%; Score 158.4; DB 1; Length 572;
Best Local Similarity 69.2%; Pred. No. 9.4e-32;
Matches 216. Conservative 0. Mismatches 06. Indels 0

QY 56 CTCTGTCTCCTCCTCCTCCTGTCCTGGGGGCTGTTGGTGTCTAGCAAGACCCTGTGCTCC 115

50 CACGCTTTTCCCTTTCTTCCCTGCTGCGGGCCCGCCAGCATGTCACCTGTTGCCCC
 116 ATGGAAGAAGCCATCAATGAGAGGATCCAGGAGTCCGCGGGCTCCCTAATATTTAGGGCA 175

Db

110 ATGSATGAAGCCATCAGCAAGAAGATCAACTACAGACTTCAGTCCCCTACTGCCAGGTGCA 169

Ov

176 ATAACCAGCATTTGGCCCCCGTGAGTAGTGCACACCCCTCACACCTCCACCTCCCGAATTGACCGC 238

2293
 2294
 2295
 2296
 2297
 2298
 2299
 2300
 2301
 2302
 2303
 2304
 2305
 2306
 2307
 2308
 2309
 2310
 2311
 2312
 2313
 2314
 2315
 2316
 2317
 2318
 2319
 2320
 2321
 2322
 2323
 2324
 2325
 2326
 2327
 2328
 2329
 2330
 2331
 2332
 2333
 2334
 2335
 2336
 2337
 2338
 2339
 2340
 2341
 2342
 2343
 2344
 2345
 2346
 2347
 2348
 2349
 2350
 2351
 2352
 2353
 2354
 2355
 2356
 2357
 2358
 2359
 2360
 2361
 2362
 2363
 2364
 2365
 2366
 2367
 2368
 2369
 2370
 2371
 2372
 2373
 2374
 2375
 2376
 2377
 2378
 2379
 2380
 2381
 2382
 2383
 2384
 2385
 2386
 2387
 2388
 2389
 2390
 2391
 2392
 2393
 2394
 2395
 2396
 2397
 2398
 2399
 2400
 2401
 2402
 2403
 2404
 2405
 2406
 2407
 2408
 2409
 2410
 2411
 2412
 2413
 2414
 2415
 2416
 2417
 2418
 2419
 2420
 2421
 2422
 2423
 2424
 2425
 2426
 2427
 2428
 2429
 2430
 2431
 2432
 2433
 2434
 2435
 2436
 2437
 2438
 2439
 2440
 2441
 2442
 2443
 2444
 2445
 2446
 2447
 2448
 2449
 2450
 2451
 2452
 2453
 2454
 2455
 2456
 2457
 2458
 2459
 2460
 2461
 2462
 2463
 2464
 2465
 2466
 2467
 2468
 2469
 2470
 2471
 2472
 2473
 2474
 2475
 2476
 2477
 2478
 2479
 2480
 2481
 2482
 2483
 2484
 2485
 2486
 2487
 2488
 2489
 2490
 2491
 2492
 2493
 2494
 2495
 2496
 2497
 2498
 2499
 2500
 2501
 2502
 2503
 2504
 2505
 2506
 2507
 2508
 2509
 2510
 2511
 2512
 2513
 2514
 2515
 2516
 2517
 2518
 2519
 2520
 2521
 2522
 2523
 2524
 2525
 2526
 2527
 2528
 2529
 2530
 2531
 2532
 2533
 2534
 2535
 2536
 2537
 2538
 2539
 2540
 2541
 2542
 2543
 2544
 2545
 2546
 2547
 2548
 2549
 2550
 2551
 2552
 2553
 2554
 2555
 2556
 2557
 2558
 2559
 2560
 2561
 2562
 2563
 2564
 2565
 2566
 2567
 2568
 2569
 2570
 2571
 2572
 2573
 2574
 2575
 2576
 2577
 2578
 2579
 2580
 2581
 2582
 2583
 2584
 2585
 2586
 2587
 2588
 2589
 2590
 2591
 2592
 2593
 2594
 2595
 2596
 2597
 2598
 2599
 2600
 2601
 2602
 2603
 2604
 2605
 2606
 2607
 2608
 2609
 2610
 2611
 2612
 2613
 2614
 2615
 2616
 2617
 2618
 2619
 2620
 2621
 2622
 2623
 2624
 2625
 2626
 2627
 2628
 2629
 2630
 2631
 2632
 2633
 2634
 2635
 2636
 2637
 2638
 2639
 2640
 2641
 2642
 2643
 2644
 2645
 2646
 2647
 2648
 2649
 2650
 2651
 2652
 2653
 2654
 2655
 2656
 2657
 2658
 2659
 2660
 2661
 2662
 2663
 2664
 2665
 2666
 2667
 2668
 2669
 2670
 2671
 2672
 2673
 2674
 2675
 2676
 2677
 2678
 2679
 2680
 2681
 2682
 2683
 2684
 2685
 2686
 2687
 2688
 2689
 2690
 2691
 2692
 2693
 2694
 2695
 2696
 2697
 2698
 2699
 2700
 2701
 2702
 2703
 2704
 2705
 2706
 2707
 2708
 2709
 2710
 2711
 2712
 2713
 2714
 2715
 2716
 2717
 2718
 2719
 2720
 2721
 2722
 2723
 2724
 2725
 2726
 2727
 2728
 2729
 2730
 2731
 2732
 2733
 2734
 2735
 2736
 2737
 2738
 2739
 2740
 2741
 2742
 2743
 2744
 2745
 2746
 2747

295
236 CCCCAGAGCTTCGCGGTACCGGCTGCACCTTGTGGCTCGGCCCTCGCTGGGATGTG 295
230 CCAGAGGCACACCGTCACTAGTCTCCTGTGGCTGTGGCTCATGGGACGTC 289

[illegible]

RESULT	5	
V84056		
ID	V84056 standard; cDNA; 560 BP.	
AC	V84056;	
DT	09-MAR-1999 (first entry)	
DE	CNA encoding a cysteine rich soluble protein designated C19.	
KW	Cysteine rich soluble protein; CRSP; C19; cell development;	
KW	mammalian immune system; antibody; abnormal proliferation; cancer;	
KW	inflammation; degeneration; regeneration; atrophy; ss.	
OS	Mus sp.	
Key		Location/Qualifiers
FH	64.408	
CDS		/*tag= a c19
FT		/end=
ET		ET

```

FT      /product= C19
FT      64..123
FT      /*tag= b
FT      124..405
FT      /*tag= c
FT      193..194
FT      /*tag= d
FT      /note= "intron present between these nucleotides in
FT      the genomic DNA"
FT      271..272
FT      /*tag= e
FT      /note= "intron present between these nucleotides in
FT      the genomic DNA"
FT
FT      WO9858061-A1.
FT      23-DEC-1998.
FT      PN
FT      18-JUN-1998;
FT      PD
FT      09-OCT-1997;
FT      PF
FT      19-JUN-1997;
FT      PR
FT      19-JUN-1997;
FT      PR
FT      19-JUN-1997;
FT      PR
FT      (SCHE ) SCHERING CORP.
FT      PA

```


FT	/number= 2
FT	1483..1571
FT	/*tag= f
FT	/number= 2
FT	1572..1690
FT	/*tag= g
FT	/number= 3
FT	1691..1792
FT	/*tag= h
FT	/number= 3
FT	1793..1993
FT	/*tag= i
FT	/number= 4
FT	1994..2437
FT	/*tag= j
FT	/number= 4
FT	2438..2588
FT	/*tag= k
FT	/number= 5
FT	2589..2708
FT	/*tag= l
FT	/number= 5
FT	2709..2796
FT	/*tag= m
FT	/number= 6
FT	2797..3074
FT	/*tag= n
FT	/number= 6
FT	3075..3132
FT	/*tag= o
FT	/number= 7
FT	3133..10631
FT	/*tag= p
FT	/number= 7
FT	10632..10726
FT	/*tag= q
FT	/number= 8
FT	10727..10879
FT	/*tag= r
FT	/number= 8
FT	10888..11008
FT	/*tag= s
FT	/number= 9
FT	11009..11099
FT	/*tag= t
FT	/number= 9
FT	11100..11186
FT	/*tag= u
FT	/number= 10
FT	11190..11820
FT	/*tag= v
FT	11723..11728
FT	/*tag= w

CC products and method can be used to treat or diagnose renal diseases and disorders selected from vesical-ureteral reflux, uni or bilateral CC pelvi-ureteral junction obstruction, multicystic renal dysplasia, renal CC agenesis, renal aplasia, hydronephrosis, Von Mayer-Rokitansky-Kuester CC disorder and bifid ureter.

SQ Sequence 11820 BP; 745 A; 1498 C; 1443 G; 683 T;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: October 1, 1999, 11:58:03 : Search time 164.82 Seconds
(without alignments)
5421.415 Million cell updates/sec

Title: US-09-099-898-1
Perfect score: 453
Sequence: 1 GTGTGCGGATTTGGTTAGC.....GGGGAGCTGGAAATAACCT 453

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	422.4	93.2	477	31	AA3111223	AA311223 EST181980
2	175.2	38.7	194	24	N41594	N41594 yw66b04.r1
3	152.4	33.6	480	38	AA796548	AA796548 vpl4f11.r
4	152.4	33.6	530	47	AI527863	AI527863 ufl3f10.y
5	152.4	33.6	535	47	AI528963	AI528963 ucl3e05.y
6	150.8	33.3	484	38	AA796118	AA796118 vo66e08.r
7	150.8	33.3	508	39	AA823170	AA823170 vw41e10.r
8	143.4	31.7	432	42	AI121371	AI121371 uc30h11.r
9	140.4	31.0	493	29	AA139082	AA139082 mrl1g09.r
10	136.2	30.1	379	26	W42057	W42057 mb16d01.r1
11	134	29.6	496	41	AI021019	AI021019 us99a04.r
12	132.8	29.3	372	26	W42069	W42069 mb16d04.r1
13	132	29.1	445	33	AA423511	AA423511 v876b03.r
14	131	28.9	371	43	AI180917	AI180917 ub77c04.r
15	129.6	28.6	511	34	AA467519	AA467519 ve01d11.r
16	128	28.3	385	44	AI286748	AI286748 ub96a04.r
17	127.4	28.1	402	28	AA097250	AA097250 mk16a08.r
18	124.2	27.4	472	38	AA795022	AA795022 vs08e05.r
19	123	27.2	453	41	AI036421	AI036421 ub64a02.r
20	120.2	26.5	520	34	AA472183	AA472183 vg98g06.r
21	104.2	23.0	227	41	AI021285	AI021285 ub07a10.r
22	98.8	21.8	271	40	AA980745	AA980745 ua45b05.s
23	96.6	21.3	309	42	AI120050	AI120050 uc26d09.r
24	96.2	21.2	422	27	W34958	W34958 mc34f05.r1
25	91.8	20.3	419	31	AA286376	AA286376 vc48c04.r
26	89.6	19.8	290	37	AA718169	AA718169 vu54h05.r
27	88.2	19.5	577	34	AA524300	AA524300 rg32q12.s
28	85.2	18.8	403	49	AV011470	AV011470 AV011470
29	79	17.4	513	37	AA692356	AA692356 vt58f01.r
30	79	17.4	407	37	AA734659	AA734659 vr90b11.r
31	78	17.2	487	37	AA711012	AA711012 vt52h07.r
32	77.8	17.2	586	36	AA615920	AA615920 vo91d01.r
33	77.8	17.2	388	37	AA691748	AA691748 vs13h05.r
34	77.4	17.1	484	37	AA711094	AA711094 vt55e11.r
35	77.2	17.0	389	30	AA273994	AA273994 vb70e06.r
36	76.6	16.9	319	49	AV021952	AV021952 AV021952
37	76.4	16.9	527	30	AA245583	AA245583 my52d07.r
38	76.2	16.8	319	39	AA823296	AA823296 vp37a09.r
39	75.8	16.7	626	34	AA518288	AA518288 vi27h06.r
40	75.6	16.7	469	37	AA711284	AA711284 vt11c12.r
41	75.4	16.6	490	37	AA689884	AA689884 vt62f12.r
42	75.2	16.6	408	35	AA592737	AA592737 vo37a09.r
43	74.2	16.4	522	36	AA638656	AA638656 vo55h09.r
44	74	16.3	524	39	AA822133	AA822133 vp35c03.r
45	73.4	16.2	354	36	AA638525	AA638525 vo54c10.r

ALIGNMENTS

RESULT 1
AA311223
LOCUS AA311223 477 bp mRNA
DEFINITION EST181980 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA311223
NID g1963551
VERSION AA311223.1 GI:1963551
KEYWORDS EST.

```

SOURCE
ORGANISM
REFERENCE
AUTHORS
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 477)
Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,
Phillips R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geoghagen N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu J.-I., Marmaros S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Saudak D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dinkes D., Feng D.-F., Ferrie A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hwang J., Li H., Meissner P.S., Olsen H.,
Raymond L., Wei X.F., Xing J., Xu C., Yu G.L., Ruben S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,
Fraser C.M. and Venter J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
On Sep 12, 1996 this sequence version replaced gi:1398072.
Other_ESTs: RHC144666
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M3 Reverse.
Location/Qualifiers
1. .477
/organism="Homo sapiens"
/db_xref="ATCC (inhost):157191"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells v"
/cell_type="T-lymphocyte"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 81 a 132 c 167 g 94 t 3 others
ORIGIN
Query Match 93.2%; Score 422.4; DB 31; Length 477;
Best Local Similarity 98.5%; Pred. No. 1.5e-88;
Matches 447; Conservative 0; Mismatches 3; Indels 4; Gaps 2;
QY 1 GTGTGCGGATTTGGTAGCTAGAGCCACCGAGAGGCGCTGCAGAGTAAGACTCTCTG 60
Db 1 GTGTGCGGATTTGGTAGCTAGAGCCACCGAGAGGCGCTGCAGAGTAAGACTCTCTG 60
QY 61 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 61 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
QY 121 AGAAGCCATCAATGAGAGGATCAGAGAGTCCGCGCTCTATATTTAGGGCAATAAG 180
Db 121 AGAAGCCATCAATGAGAGGATCAGAGAGTCCGCGCTCTATATTTAGGGCAATAAG 180
QY 181 CAGCATTTGGCTGAGTGCAGAGCGTCACCTCCAGGGGGGACCTGCTACTTCCGCCCG 240
Db 181 CAGCATTTGGCTGAGTGCAGAGCGTCACCTCCAGGGGGGACCTGCTACTTCCGCCCG 240
241 AGGCTTCGCCGT-CACCGGCTGCACCTGTGGCTCCGCTGTGGCTGTGGATGTGCGCG 299

```

```

Db 241 AGGCTTCGCCGTNCACCGGCTGCACCTGTGGCTCCGCTGTGGCTGTGGATGTGCGCG 300
QY 300 CCAGAGACCATGTCACCTGCGAGTGGCGGCGCATGACTGGACGGAGCGCGTCTGTCTC 359
Db 301 CCAGAGACCATGTCACCTGCGAGTGGCGGCGCATGACTGGACGGAGCGCGTCTGTCTC 360
QY 360 GTGTGAGCCCTGAGTTCGCGGCGCGCGCTGCACAGCGCGGCGGCGGCTCCAGGT 419
Db 361 GTGTGAGCCCTGAGTTCGCGGCGCGCGTGC---NACAGCGCGGCGGCGGCTCCAGGT 417
QY 420 CCGGAGGGTTGCGGGGAGCTGGAATAAACCT 453
Db 418 CCGGAGGGTTGCGGGGAGCTGGAATAAACCT 451
RESULT 2
N41594 194 bp mRNA EST 24-JAN-1996
LOCUS yw66h04.r1 Soares_placenta_8to9weeks_2NbHP8to9w Homo sapiens cDNA.
DEFINITION clone IMAGE:257239 5', mRNA sequence.
ACCESSION N41594
NID 91185625
VERSION N41594.1 GI:1165625
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194)
AUTHORS Hollman L., Clark N., Dubuque T., Elliston K., Hawkins M.,
Hollman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
Trevaskis E., Waterston R., Williamson A., Wohlmann P. and
Wilson R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 9, 1995 this sequence version replaced gi:802396.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 174
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 174.
FEATURES
source
1. .194
/organism="Homo sapiens"
/db_xref="GDB:3886849"
/db_xref="taxon:9606"
/clone="IMAGE:257239"
/dev_stage="two placenta: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGCATGTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 30 a 60 c 62 g 39 t 3 others
ORIGIN

```

T 3'); double-stranded cDNA was ligated to Eco RI					
adaptors (Pharmacia), digested with Not I and cloned into					
the Not I and Eco RI sites of the modified p7T3 vector.					
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library					
constructed and normalized by Bento Soares and M.Fatima					
Bonaldo.*					
BASE COUNT	107 a	133 c	121 g	119 t	
ORIGIN					
Query Match	33.6%	Score 152.4;	DB 38;	Length 480;	
Best Local Similarity	67.8%;	Pred. No. 3.6e-26;			
Matches 213;	Conservative	0;	Mismatches 101;	Indels	0; Gaps 0;
QY	54	CTCTGTCTCCRCCTCCTCCCCTGTCCTCGGGGCTTGTTGGTGTCTAGCAAGACCCCTGTGCT	113		
Db	87	CCCTCCTTTTCCCTTTTCTTCTTGTCCCTGAAGCTGCTGGGCTCCAGCATGCCACTGTGTGC	146		
QY	114	CCATGGAAGAAGCATCAATGACGAGGATCCAGGAGTTCGGCGGCTCCCTAATAATTAGGG	173		
Db	147	CCATCGATGAAGCATCGACAAGAAGATCAACAAGAACTTCAACTCCCTGTTCCTCAAATG	206		
QY	174	CAATAAGCACATTTGGCTGGAGTGCCAGAGCGTCACTCCAGGGGGACCTGGCTACTT	233		
Db	207	CATAAAGAACATTTGGCTTAATTTCTGGACAGTCTCCTCCAGAGGAAGTTGGCCTCT	266		
QY	234	GCCCCGAGGCTTCGCCGTCACCGGCTGCACATTGTGGCTCCGCTGTGGCTCGTGGGATG	293		
Db	267	GCCAGAAGGCACAGCAGTCTTAGAGCTGCTCCTGTGGCTCTGCCTGTGGCTGTGGGACA	326		
QY	294	TGGCGCCGAGACACATGTCATCCAGTCGCGCGGCGATGGATGGACCGGAGCGCGCT	353		
Db	327	TTGCTGAAGAAAAAGTGTGCTCACTGCCAGTGTCCAAGTAGACTGGACAGCAGCCGCT	386		
QY	354	GCTGTCGTGTGCAG	367		
Db	387	GCTGTAAGCTGCAG	400		
RESULT	4				
AI527663					
LOCUS	AI527663	530 bp	mRNA	EST	18-MAR-1999
DEFINITION	uf3f10.y1 Soares mouse mammary gland NLMG Mus musculus cDNA clone IMAGE:1511275 5', mRNA sequence.				
ACCESSION	AI527663				
NID	g4441798				
VERSION	AI527663.1 GI:4441798				
KEYWORDS	Est.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 530)				
	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.				
TITLE	The WashU-NCI Mouse EST Project 1999				
JOURNAL	Unpublished (1999)				
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3187318. Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:938127 Seq primer: -40RP from Gibco High quality sequence stop: 481.				

[illegible]

Best Local Similarity 67.5%; Pred. No. 8.6e-26;
Matches 212; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
QY 54 CTCCTGTCTCTCTCCCTCCCTGCTGCTGGGGTGTGGTGTCTAGCAAGACCTGTGCT 113
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 CCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 CCATGAAGAAGCCATCAATGAGAGATCCAGAGGTCCCGGCTCCCTAAATATTTAGG 173
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 CCATGATGAAGCCATCGACAAGATCAACAAGACTTCAACTCCCTGTTTCCAAATG 177
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 CAATAGCAGCATTCCTTGGAGTCCAGAGCGTCACTCCAGGGGGGACCTGGCTACTT 233
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 CAATAAAGAACATTCGCTTAAATGCTGCACAGTCTCTCCAGAGGAAGTTGGCCCTCT 237
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 GCGCCGAGGCTTCGCGCTCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 293
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 GCCCAGAGGACAGAGCTTGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 TCGCGCCGAGACCATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 353
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 TTCTGTGAAGAAAAGTGTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 357
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 354 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 358 GCTGTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 371
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 8

AA121371
LOCUS uc30h11.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
DEFINITION IMAGE:1399557 5', mRNA sequence.
ACCESSION AA121371
NID 93521695
VERSION AA121371.1 GI:3521695
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 9, 1998 this sequence version replaced gi:937777.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:911273
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 426.
Location/Qualifiers

FEATURES

source

1. .432
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1399557"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"

/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site:1; Not I;
Site:2; Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dr) primer [5',
TGTTCACCAATCGAAGTGGAGCGGCGGGAATGTTTTTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonalido."

BASE COUNT 99 a 117 c 110 g 106 t
ORIGIN

Query Match 31.7%; Score 143.4; DB 42; Length 432;
Best Local Similarity 66.9%; Pred. No. 4.2e-24;
Matches 204; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
QY 54 CTCCTGTCTCTCTCCCTCCCTGCTGCTGGGGTGTGGTGTCTAGCAAGACCTGTGCT 113
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 CCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 144
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 CCATGAAGAAGCCATCAATGAGAGATCCAGAGGTTCGCGGCTCCCTAAATATTTAGG 173
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145 CCATGATGAAGCCATCGACAAGATCAACAAGACTTCAACTCCCTGTTTCCAAATG 204
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 CAATAAGCAGCATTCGCTGAGTCCAGAGCGTCACTCCAGGGGGGACCTGGCTACTT 233
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 CAATAAAGAACATTCGCTTAAATGCTGAGACAGTCTCTCCAGAGGAAGTTGGCCCT 264
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 GCGCCGAGGCTTCGCGCTCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 293
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 GCCCAGAGGACAGCAGTCTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 324
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 TCGCGCCGAGACCATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 353
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 TTCGTAAGAAAAGTGTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 384
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 354 GCTGT 358
  |||||
Db 385 CGTGT 389
```

RESULT 9

AA139082
LOCUS mrl1q09.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:597184
DEFINITION 5', mRNA sequence.
ACCESSION AA139082
NID 91701283
VERSION AA139082.1 GI:1701283
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397652.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

AIO21019		496 bp	mRNA	EST	16-JUN-1998
ua9ba04.r1	Soares mouse mammary gland NbMMG Mus musculus cDNA clone				
IMAGE:	I365582 5'	MRA sequence.			
AIO21019	NID	g3235355			
AIO21019.1	GI:3235355				
EST.	house mouse.				
Mus musculus	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	(bases 1 to 496)				
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	The WashU-HMI Mouse EST Project Unpublished (1996) On Jan 19, 1998 this sequence version replaced gi:2151315.				
TITLE JOURNAL COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer:-'28ml3 rev2 Et from Amersham High quality sequence stop: 458. Location/Qualifiers l..496 /organism='Mus musculus' /strain='C57BL/6J' /db_xref='taxon:10090' /map='21q' /clone='IMAGE:I365582' /clone_lib='Soares mouse mammary gland NbMMG' /sex='male' /tissue_type='mammary gland' /dev_stage='4 weeks' /lab_host='DH10B' /note='Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I-- oligo(dt) primer [5' TGTTACCAATCTCAAGTCGGCGGCCGCAGTGATTGTCTTTTTTTTTTTTTTTTTTTT T 3']; adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo.'				
BASE COUNT ORIGIN	109 a	140 C	124 G	123 t	
Query Match Best Local Similarity Matches	29.6%; Score 134; DB 41; Length 496; 67.08; Pred No. 6.5e-22; Conservative	0;	Mismatches 100;	Indels 1;	Gaps 1;
QY Db QY Db QY Db QY	54 CTCTCTCTCCCTCCTCCCTCCCTGCTGCGGGCTGTGGTGTCTTAGCACGAACCCTGTGCT 113 87 CCCTCCCTTTCTCTCTCTCTCTCTGCTGAACCTGCTGGGCTCCA-CATGCCACTGTGTC 145 114 CCATGGAAGAAGCATCAATCAGAGGATCCAGGAGGTCCGCCGCTCCCTATAATTAGGG 173 146 CCATCGATGAAGCATCGACAAGAAGATCAAACAAGACTTCAACTCCCTGTTTCCAATG 205 174 CAATAAAGCAGATTGGCCTGGAGTGCACAGACGCTCACCTCCAGGGGGGACCTGGCTACTT 233				

[illegible]

BASE COUNT
ORIGIN

102 a 120 c 113 g ~ 110 t

/note="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site.1:
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(du) primer [5',
TGTTACCAATCTGAGTGGAGCGGCCGGAATGTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

Query Match	29.1%;	Score 132;	DB 33;	Length 445;
Best Local Similarity	65.9%;	Pred. No.	1.8e-21;	
Matches 207;	Conservative 0;	Mismatches 105;	Indels 2;	Gaps 1;
<hr/>				
QY	54	CNCTGTGTCCTCCNCCCTCCCTGTGCTGGGGCTGTGTGGTCTACAGACCCTGTGCT	113	
Dd				
Dd	91	CCCTCCTTTCTTTCTTTCTTGTCCCCTGAACGTGTGGCTCCAGCATGCCACTGTGC	150	
QY	114	CCATGGAAGAAGCCATCAATGAGAGATCCAGGAGTGCGCGGCTCCCTAAATTATAGGG	173	
Dd				
Dd	151	CCATCATGATGCCAATCGACAAGATCAAACAAGACTTCACTCCCTGTTCCAATG	210	
QY	174	CAATAAGCAGCATTTGCCCTGGAGTCCAGAGCGTCACTCCAGGGGGACCTGGCTACTT	233	
Dd				
Dd	211	CAATAAGAACAATTTGGCTTTAAAATTTGCTGGACAGTCTCCTCCAGAGGAAGTTGGCCTCCT	270	
QY	234	GCCCCGAGGCTTCGCCGTACCCGGTGCACATGTGTGGCTCCGCTCTGTGCTGGGATG	293	
Dd				
Dd	271	GCCCAGAAGCACAGCAGTCTTAGCTGTCTCTGTGGCTCTGC--TGTGGTCTGTGGGACA	328	
QY	294	TGCGGCCCGAGACCCAGATGTCAGTGCAGTGC CGCGGCATGGACTGGACCGGAGCGCGT	353	
Dd				
Dd	329	TTCGTGAAGAAAAGTGTCTACTGCCATGTGC AAGGATAGACTGGACAGACGCCGCT	388	
QY	354	GCTGTCGTGTGCAG	367	
Dd				
Dd	389	GCTGTAAAGCTGCAG	402	

RESULT 14
 AII180917
 LOCUS
 DEFINITION
 ACCESSION
 NID
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

AII180917 371 bp mRNA EST 08-OCT-1998
 ub77c04.r1 Soares mouse mammary gland NMDMG Mus musculus cDNA clone
 IMAGE:1383750 5', mRNA sequence.
 AII180917
 g3731555
 AII180917.1 GI:3731555
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 371)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 On Sep 12, 1996 this sequence version replaced gi:1407470.

Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 11:58:04 ; Search time 144.5 Seconds
(without alignments)
9970.105 Million cell updates/sec

Title: US-09-099-898-1
Perfect score: 453
Sequence: 1 GTCTGCCGATTTGGTTAGC.....GGGAGCTGGAATAAACCT 453

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmb1.*

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pil.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: gb_htg1.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: em_ba1.*
- 39: em_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
1	43.8	9.7	981 12	RNCKERIMP X54806 R.norvegicu

2	43.4	9.6	41055	1	SC5G4
3	41.6	9.2	40883	11	AC005776
4	40.8	9.0	39441	11	AC005565
5	40.8	9.0	5180	12	RNU48596
6	40.6	9.0	47852	1	MTV023
7	40.2	8.9	47852	1	MTV023
8	40	8.8	5880	17	HSE4
9	39.8	8.8	7082	1	SF008223
10	39.8	8.8	1320	3	CVCMYC
11	39.4	8.7	806	7	HVU49482
12	39.4	8.7	2578	12	MUSHOXA
13	39.2	8.7	2157	4	XELB1AA
14	39.2	8.7	3252	10	HSTAF113
15	39.2	8.7	42189	11	AC005258
16	39	8.6	129354	11	AC004080
17	39	8.6	2691	11	AF040714
18	39	8.6	3637	12	AF045017
19	39	8.6	3579	12	MMAJ3298
20	39	8.6	5123	17	HEPVIE
21	39	8.6	8438	17	SHILLT
22	38.8	8.6	907	7	HVGRP3
23	38.8	8.6	170030	11	AC005829
24	38.6	8.5	4246	11	AF055376
25	38.6	8.5	2145	11	AF055377
26	38.6	8.5	2579	12	MUSMDR2A
27	38.4	8.5	3975	1	AMPEPSYNT
28	38.4	8.5	40790	1	MTCY493
29	38.2	8.4	2427	9	BABINTB5GL
30	38.2	8.4	244254	10	HSAC001228
31	38.2	8.4	4000	12	HSAC001228
32	38	8.4	67200	1	MTV017
33	37.8	8.3	36583	1	SC5H1
34	37.8	8.3	785	8	AF010580
35	37.8	8.3	5721	12	RNCA171
36	37.6	8.3	164296	34	HSAC000380
37	37.4	8.3	84364	9	HS222J7
38	37.4	8.3	35465	11	AC005525
39	37.4	8.3	110000	34	HSDJ689N3_1
40	37.4	8.3	3372	42	AF065164
41	37.2	8.2	7515	11	HSU26644
42	37	8.2	43632	1	SC8A5
43	36.8	8.1	2028	8	SSI132828
44	36.8	8.1	5225	10	HSU06154
45	36.8	8.1	1756	10	HUMD2A

ALIGNMENTS

RESULT	1
RNCKERIMP	
LOCUS	R.norvegicus mRNA for cytokeratin type I (3' end).
DEFINITION	RNCKERIMP 981 bp mRNA ROD 30-APR-1992
ACCESSION	X54806
NID	955955
VERSION	X54806.1 GI:55955
KEYWORDS	cytokeratin; cytokeratin type I.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE	1 (bases 1 to 981)
JOURNAL	Redfern,C.P.F.
REFERENCE	Direct Submission
AUTHORS	Submitted (26-SEP-1990) Redfern C.P.F., Univ. of Newcastle upon
TITLE	Tyne, Medical Mol. Biology Group, Dept. of Dermatology, 4th Floor
JOURNAL	Cookson Bldg. Medical School, Framlington Place, Newcastle upon
REFERENCE	Tyne NE2 4HH, UK
AUTHORS	2 (bases 1 to 981)
TITLE	Redfern,C.P.F. and Allen,G.
JOURNAL	Sequence variation (between species) at the carboxyterminal domain
	of cytokeratin-1
	Unpublished

- AL031317 Streptomy
- AC005776 Homo sapi
- AC005565 Homo sapi
- U48596 Rattus norv
- AL022022 Mycobacte
- AL022022 Mycobacte
- D14486 Equine herp
- U08223 Streptomyce
- X95367 C.familiari
- U49482 Hordeum vul
- L08757 Mus musculu
- M63663 X.borealais
- Y11354 H.sapiens m
- AC005258 Homo sapi
- AC004080 Homo sapi
- AF040714 Homo sapi
- AF045017 Mus muscu
- AJ223298 Mus muscu
- X15120 Pseudorabie
- M57505 Pseudorabie
- 248624 H.vulgare m
- AC005829 Homo sapi
- AF055376 Homo sapi
- AF055377 Homo sapi
- M74151 Mouse p-gly
- X97860 Mycolatops
- 295844 Mycobacteri
- L12231 Yellow babo
- AC001228 244Kb Con
- M88299 Mouse brain
- AL021897 Mycobacte
- AL049863 Streptomy
- AF010580 Oryza sat
- Z78279 R.norvegicu
- AC000380 Homo sapi
- Z98885 Human DNA s
- AC005525 Homo sapi
- Continuation (2 of
- AF065164 Homo sapi
- U26844 Human fatty
- AL049485 Streptomy
- AJ132828 Spermatoz
- U06154 Human clone
- M30625 Human dopam


```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS       Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 40883)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V.,
Burkhardt-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Kronmiller,B., Arellano,A., Saunders,C., Ov,D., Nolan,M., Trong,S.,
Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 1.5 Mb OLFER-rich region in 19p13.1
Unpublished
2 (bases 1 to 40883)
Lamerdin,J.E.
Direct Submission
Submitted (05-OCT-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from p telomere to centromere. Cosmid
R28550 overlaps cosmid R28194 (AC003111) to the left from bases 1
to 504 of this accession, and is separated from cosmid R33907 to
the right by a sequence gap of <2 kb. Additional chromosome 19 map
and sequence information are available at:
http://www-bio.lnl.gov/db/rp/genome/genome.html.
FEATURES     Location/Qualifiers
source       1..40883
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="R28550"
             /chromosome="19"
             /map="19p13.1 between OLFER and D19S885"
             /cell_line="5HL2-B"
             /clone_lib="LL19NC02 F chromosome 19 specific cosmid
             library"
             /note="Cosmid library constructed at LNL from flow-sorted
             chromosomes from human-hamster hybrid 5HL2-B, which
             carries chromosome 19 as its only human chromosome."
repeat_region 1..245
             /rpt_family="AluSq"
repeat_region 246..442
             /rpt_family="LIMB8"
repeat_region 443..754
             /rpt_family="AluSq"
repeat_region 770..963
             /rpt_family="LIMB1"
repeat_region 1161..1282
             /rpt_family="AluSq/x"
repeat_region 1286..1585
             /rpt_family="AluSq"
repeat_region 1588..1880
             /rpt_family="AluSq"
repeat_region 1891..2012
             /rpt_family="AluJb"
repeat_region 2024..2326
             /rpt_family="AluY"
repeat_region 2327..2616
             /rpt_family="AluSq"
repeat_region 2776..2841
             /rpt_family="LINE2"
repeat_region 3495..3622
             /rpt_family="LINE2"
repeat_region complement(3667..3766)
             /rpt_family="MER3"
misc_feature complement(4287..4396)
             /note="predicted exon, program: grill2exons_human_1.3,
             frame: 2, quality: good, score: 52.000"
repeat_region 4507..4787
             /rpt_family="AluSq"
repeat_region 4808..5109
             /rpt_family="AluY"
repeat_region 5283..5545
             /rpt_family="AluSq"
6064..6207
             /rpt_family="MIR"
repeat_region 6537..6577
             /rpt_family="(TGG)n"
repeat_region 8122..8401
             /rpt_family="AluJb"
repeat_region 8402..8565
             /rpt_family="AluJb"
repeat_region 8585..8889
             /rpt_family="AluJb"
repeat_region 9886..10177
             /rpt_family="AluSx"
repeat_region 10179..10251
             /rpt_family="(GAAA)n"
repeat_region complement(11205..11323)
             /rpt_family="AluY"
repeat_region 11363..11494
             /rpt_family="AluY"
repeat_region 11506..11820
             /rpt_family="AluY"
repeat_region 11821..12003
             /rpt_family="AluYa5"
repeat_region 12022..12128
             /rpt_family="FLAM_A"
repeat_region complement(12200..12259)
             /rpt_family="MIR"
repeat_region complement(12910..13002)
             /rpt_family="MIR"
misc_feature 13871..14036
             /note="predicted exon, program: grill2exons_human_1.3,
             frame: 1, quality: excellent, score: 83.000"
repeat_region complement(14095..14209)
             /rpt_family="MIR"
repeat_region 14662..14951
             /rpt_family="AluSp"
repeat_region complement(15085..15108)
             /rpt_family="AT-rich"
repeat_region 15109..15431
             /rpt_family="AluSx"
misc_feature 15662..15770
             /note="predicted exon, program: grill2exons_human_1.3,
             frame: 1, quality: excellent, score: 80.000"
repeat_region 16107..16190
             /rpt_family="LIMB8"
repeat_region 16191..16497
             /rpt_family="AluSx"
repeat_region 16498..16937
             /rpt_family="LIMB8"
repeat_region complement(17602..17708)
             /rpt_family="LINE2"
repeat_region complement(17709..18005)
             /rpt_family="AluSx"
repeat_region 18393..18690
             /rpt_family="AluSc"
misc_feature 18706..18968
             /note="BLASTN similarity to 741380 (105..368); match:
             0.99, score: 3.8e-100; database searched: est;
             phib12_19/1rv Homo sapiens cDNA clone phib12_19/1rv."
repeat_region 19418..19703
             /rpt_family="AluSx"
repeat_region complement(20667..20801)
             /rpt_family="GC-rich"
repeat_region complement(20802..20829)
             /rpt_family="(GGA)n"
repeat_region complement(20871..20955)
             /rpt_family="(CGGG)n"
repeat_region complement(20982..21134)
             /rpt_family="GC-rich"
repeat_region complement(21237..21352)
             /rpt_family="(CGGG)n"
repeat_region complement(21349..21470)
             /rpt_family="(CGG)n"
repeat_region 21482..21559

```


[illegible]

Query Match	9.0%	Score 40.8	DB 12	Length 5180;
Best Local Similarity	50.0%	Pred. No.	4.5;	
Matches 102;	Conservative	0;	Mismatches 102;	Indels 0; Gaps 0;

QY	201	AGAGCGTCACTCCAGGGGACCTGGCTACTTTGCCCGCAGGCTTCGCCGTACCGGCT	260
Dd	607	AGAGCTCCGCGCGCGCGCGCGCTCGGGCTCGCGCGCGCGCGCGCGCGGAATCCGCAC	548
QY	261	GCACTTGTCCTCGCGCTGTGGATGTGGCGCGCGAGACCACATGTCACTGCC	320
Dd	547	GACGAGCGCATGCGCGCGCGCGCGCGCGCATTTCTCGCGGCTACATTCGCTCGTG	488
QY	321	AGTCGCGGCGATGAGTGTGACCGGAGCGCGCTGTCTGTCTGTGTCAGCCCTGAGGTCGCG	380
Dd	487	CCTGGGCGGAGGAGGGGCGGGAGAGCGCGGGGCGGCGGAGCGCGCGCGCGCGCG	428
QY	381	CGAGCGCGTGTACAGCGCGCGCGCG	404
Dd	427	CGTGGCGCGAGTGTGCGCGCGCGCG	404

RESULT	6
MTV023	
LOCUS	47852 bp DNA BCT 17-JUN-1998
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 148/162..
ACCESSION	AL022022 AL123456
NID	93261554
VERSION	AL022022.1 GI:3261554
KEYWORDS	
SOURCE	Mycobacterium tuberculosis.
ORGANISM	Mycobacterium tuberculosis. Bacteria: Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

REFERENCE	1 (bases 1 to 47852)
AUTHORS	Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E., Tekala,E., Badcock,K., Basham,D., Brown,D., Chillingworth,I., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrell,B.G.
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
JOURNAL	Nature 393 (6685), 537-544 (1998)
MEDLINE	98295987
REMARK	Erratum:[published erratum appears in Nature 1998 Nov 12;396(6707):190]] 2 (bases 1 to 47852) Parkhill,J.
REFERENCE	
AUTHORS	
TITLE	Direct Submission
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2924430. Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in tParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome
COMMENT	

BASE COUNT
ORIGIN

binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

Location/Qualifiers

1. 47852

source

/organism="Mycobacterium tuberculosis"

/strain="H37Rv"

/db_xref="taxon:1773"

<1. 18

source

/organism="Mycobacterium tuberculosis"

/strain="H37Rv"

/db_xref="taxon:1773"

/clone="Y13E12"

complement(1. 1695)

CDS

/gene="Rv3494c"

/note="Rv3494c. (MTV023.01c), len: 564. Unknown protein similar to several Mycobacterium tuberculosis proteins e.g. MTC128.14 (515 aa), MTC19H5.28c (516 aa) and (MTV051.09). Has hydrophobic stretch, possibly signal peptide at N-terminus. FASTA scores: 297050|MTC128.14 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 297182|MTC19H5.28 (516 aa) opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in 555 aa overlap. TBPase score is 0.897"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv3494c"

/protein_id="CAAL7731.1"

/db_xref="PID:el254622"

/db_xref="PID:g2924431"

/db_xref="GI:2924431"

/db_xref="SPTREMBL:O53539"

/translation="MIDRLAKIQLSIFAVITVITLSVMAIFVLRPLPATEFGITGVSA
DFVAGGLYKKNATYRGVANGRVESVGLNPNGTVAHRLNLSGTAIPSNVATVRSVS
AIGQYIDLVPEPSSKLRNGFRIQONRIGQVADLLRQAETLLSGDGLTRLE
LLHEAFATNGAPGELARLIESARLLVDEANAYPQVSQLIDQAGFFLQAQIRAGGDI
KSLADGARFTWQLRAADPRDLADAPDAIDEANTAFSGIRSPFPALAAASLANLGR
VGVYHKIEQLLVVPELAFRAIITISAGVPODEGAKLDFKIDLDHPPPCMTFLPPP
LVRSAPDESVEIPRDMCKTAQNDPSVTRGARNYCOEEFGKRAPTVOLCRDPRGYV
PVGTNPWGPPPIPYTEVTDGRNLIIPNKFYIPGADPDGCVIGVPPPGQVAGPG
FAPHPQAPAPPNDNGPPPTSWMPGYPPEPPQVYPATIPPPPPPEGGPPPGP
APGPOQASGPAYTIYDLSGAFADPAGGTGIFAFGMTGASSAENWVDLMRDPRL"

complement(1. 1695)

/gene="Rv3494c"

<1. 47852

misc_feature

/note="fragment designated v023. Does not represent a physical clone"

complement(1703. 1707)

RBS

/note="possible RBS for Rv3494c"

gene

complement(1706. 2860)

gene

/gene="lprN"

CDS

complement(1706. 2860)

/gene="lprN"

/note="Rv3495c. (MTV023.02c), len: 384. lprN, similar to Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and MTCY19H5.29 (402 aa) and (MTV051.08). Probably lipoprotein, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. FASTA scores: gp|297050|MTC128.13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and 297182|MTCY19H5.29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TBPase score is 0.897"

/codon_start=1

/transl_table=11

/product="lprN"

/protein_id="CAAL7732.1"

/db_xref="PID:el254623"

/db_xref="PID:g2924432"

/db_xref="GI:2924432"

/db_xref="SPTREMBL:O53540"

/translation="NNRIWLRAIILTASSALLAGCQFGGLNSILPLPGTAGHGAYSV

TVEMADVATLPQNSPVMVDDVTGVSAGIVAVORPDGSFYAAVKLDDLKNNLLPANAV
AKVSTSLIGSLHVELAPTDPRPTGRLVDGSRITTEANTDRPTTTEVEFSALGVVYVK
GNVGALEIIDEITHQAVAGROAFVNLPRLAELTAGLNQVHDIIDALDGLNRVSAI
LARDKDNLGRALDTPDAVRLVQNRDHIYDAFAALKRLTMTVSHVLAETKVDGDDL
KDLYSIVALKNDDRDKDFVTSLLQLLTFFFPFGIKQAVRGDYLNVTTTDLTLRRIGE
TFTTAYFDPNNAHDEILNPPDFLIGELANLSGQAADPFKIPPGTASGQ"

complement(2798. 2830)

/gene="lprN"

/note="PS00013 prokaryotic membrane lipoprotein lipid attachment site"

complement(2857. 4212)

gene

/gene="Rv3496c"

CDS

complement(2857. 4212)

/gene="Rv3496c"

/note="Rv3496c. (MTV023.03c), len: 451. Unknown but similar to Mycobacterium tuberculosis proteins MTC128.12 (530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic region at N-terminus. FASTA scores: 297050|MTC128.12 (530 aa) opt: 838 z-score: 711.0 E(): 4.3e-32; 35.1% identity in 473 aa overlap; and 297182|MTCY19H5.30 (508 aa) opt: 821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap. TBPase score is 0.891"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv3496c"

/protein_id="CAAL7733.1"

/db_xref="PID:el254624"

/db_xref="PID:g2924433"

/db_xref="GI:2924433"

/db_xref="SPTREMBL:O53541"

/translation="NMGRVAMLTGSRGLRYATVIALVAALVGVVYLSSTGNKRTIVG
YFTSAGLVPGDQVRVLGVGEIDMPERSSDVKITMSVDKVPDVQVAVIMSPN
LYAARFIQLTPYTGAVLPDNGRIDLDRATVPVEVDEKGLTRLAADLSPAGELQ
GPELQAINGAADTLDNGDSLHNLRELQAQVAGRLGDSRGDIFGTVKNLQVLVDALSE
SDEQIVQAGHAVASQVLSLANSANLDTGLNLQALSDIRGFNRNNSTLIETVQL
NDFATLSDQSENIQVQLHVAGGITNFYNYIDPAQGLTISFNANVQFICGG
SPTDAGSAPADPYRRAEICRERGLVRLTVNYPIMFHLPLNTITAVKGQIYDTP
ATEAKSETVPPELTWVPGGAPVGNPADLSLLVPPAPGAPAPGAPGCGEGGG
G"

complement(2867. 2871)

RBS

/gene="Rv3496c"

/note="possible RBS for Rv3496c"

gene

complement(4209. 5282)

gene

/gene="Rv3497c"

CDS

complement(4209. 5282)

/gene="Rv3497c"

/note="Rv3497c. (MTV023.04c), len: 357. Unknown but similar to Mycobacterium tuberculosis proteins MTCY19H5.31 (481 aa), MTC128.11 (515 aa) and MTV051.06. Hydrophobic region at N-terminus. FASTA scores: 297182|MTCY19H5.31 (481 aa) opt: 611 z-score: 667.7 E(): 1.1e-29; 32.5% identity in 332aa overlap; and 297050|MTC128.11 (515 aa) opt: 587 z-score: 641.2 E(): 3.3e-28; 30.1% identity in 335 aa overlap. TBPase score is 0.889"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv3497c"

/protein_id="CAAL7734.1"

/db_xref="PID:el254625"

/db_xref="PID:g2924434"

/db_xref="GI:2924434"

/db_xref="SPTREMBL:O53542"

/translation="MLNRKPPSKHERDPLRTGIFGLVLVTCVLLIAFGYSGLFWPQGG
KYDAYFTDAGGITPGNSVYSLGKVGVASVAGNSAKVTSVDRSIYVQGSALAA
IRNTDTLGRSIAVSPAGSKTTIPLSRITTTPTLNGVLQDLGRANDLNRQFEQV
LNVFTQALHDATPOVRGAVDGLTSLRNLNRDEALQGLLAHAKSVTSVLSRAEQVN
KLVEDGNOLFAALDAPRAALSALISGIDDDVAQISGFVADNRKFCGPALEKLVLAN
LNERDYTEALKRRLPTATYTGVEVVGSGFGFNWVYSVLPGPLVATVFDLVFPQGLK
PDSLADYLRGFTOERWIIRPKSP"

complement(5272. 6324)

gene

/gene="Rv3498c"

CDS

complement(5272. 6324)

/gene="Rv3498c"


```
0: 33.6% identity in 363 aa overlap; and
297182|MTC19H5_29 (402 aa) opt: 572 z-score: 667.9 E():
1.le-29; 31.8% identity in 362 aa overlap. Tbpase
score is 0.897"
/codon_start=1
/transl_table=11
/product="lprn"
/protein_id="CAAL7732.1"
/db_xref="PID:el254623"
/db_xref="PID:g2924432"
/db_xref="GI:2924432"
/db_xref="SPTREMBL:O53540"
/translation="MNRILWRAIIITASSALLAGCQFGLNSLPLPGTAGHGEAYS
VEMADVATLQNSVPMVDDVTYVAGVAVQVRDGSFYAARLKLDDKNVLLPANAV
AKVQSTLGLSHLELAPPTDRPTGRLVDGSRTEANTDREPTTEVFESALGVVYVK
NARGALEIIDIETHQAVROAOFVNLPRLAELTAGLNROVHDIIDALDGLNVSAL
LARDKDNIGRALDITLPDARVNLQNRDHVDAFAALKRLTWTSVLAETKVDGFEDL
KDLISVAKALNDKDFVTSLOLLTFEFPNFGKQAVRGDLYLVFTTFDLRLRIGE
TFFTAYDFDPNNAHDEILNPPDFLIGELANLSGOAADPFKIPPGTASQ"
complement(2798..2830)
/gene="lprn"
/feature="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(2857..4212)
/gene="Rv3496c"
complement(2857..4212)
/gene="Rv3496c"
/note="Rv3496c. (MTV023.03c), len: 451. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.12
(530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic
region at N-terminus. FASTA scores: 297050|MTC128.12 (530
aa) opt:838 z-score: 711.0 E(): 4.3e-32; 35.1% identity
in 473 aa overlap; and 297182|MTCY19H5.30 (508 aa) opt:
821 z-score:697.0 E(): 2.6e-31; 35.1% identity in 453 aa
overlap. Tbpase score is 0.891"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3496c"
/protein_id="CAAL7733.1"
/db_xref="PID:el254624"
/db_xref="PID:g2924433"
/db_xref="GI:2924433"
/db_xref="SPTREMBL:O53541"
/translation="MMGRVAMLTGSRGLRYATVIALVAGVGVVLSSTGKRTIVG
VETSAVLPGDOVLGVPPVGEIDIEPRSSDVKTWSVKDVQVQVAVISPN
LVAAPIQLTPVTGCVLPPNGRIDLRDTPVPEWDEKGLRLAIDLSPAGELQ
GPLGAAIINQADLDGNGSLHNLRELQAQVAGRLGDSRGDIFGVKNLVLDVLSL
SDEQIVFAGHVSYSQVLADSSANLDQTLGLNQLSDIRGLRENNSTLTETYNQL
NDFQTLSDSENIEQVLHVAGPGITNFYNYDPAQGLNLGLISIPNFPANPVQICGG
SFDTAAGPSADPYRRAEICRERLGPVLRRLTVNPPIMFPLNTITAYKGOIIVDTP
ATEAKSEVPVPLTWTVPAGGAPVGNPDLQSLVPPAPGAPAPAGPAGGEGGG
G"
complement(2867..2871)
/gene="Rv3496c"
/note="possible RBS for Rv3495c"
complement(4209..5282)
/gene="Rv3497c"
complement(4209..5282)
/gene="Rv3497c"
/note="Rv3497c. (MTV023.04c), len: 357. Unknown but
similar to Mycobacterium tuberculosis proteins MTCY19H5.31
(481 aa), MTC128.11 (515 aa) and MTV051.06. Hydrophobic
region at N-terminus. FASTA scores: 297182|MTCY19H5.31
(481 aa) opt:611 z-score: 667.7 E(): 1.1e-29; 32.5%
identity in 332aa overlap; and 297050|MTC128.11 (515 aa)
opt: 587 z-score:641.2 E(): 3.3e-28; 30.1% identity in
335 aa overlap. Tbpase score is 0.889"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3497c"
/protein_id="CAAL7734.1"
/db_xref="PID:el254625"
/db_xref="PID:g2924434"
```

```
/db_xref="GI:2924434"
/db_xref="SPTREMBL:O53542"
/translation="MLNRKPSKHERDPLRTGIFGLVLCVIVLVIAFGYSGLPFWQOG
KTYDAYFTDAGGITPGNSVYSGLVKAVSAVSLAGNSAKVTFSDRSIVVGQSLAA
IRTDNIIIGERSIAVSPAGSKSTIPLSTPTTYTLNGVQLDGRNANDLNRPFQQA
LNVFTQALHDATPQVAGVDTLSLRNALNRDEALQGLLAHAKSVLSLSRAEQVN
KLVEDGNQLFAALDARRAALSALISIGFVADNAQISGFVADNRKEFGPALSKNLVLAN
LNRRDYITEALKRLPTATTGTEVYVGGPFGFNVSYSVLPGLVATVDFLVPQGLK
PDSLADYLRGFTOERWIIIRPKSP"
complement(5272..5324)
/gene="Rv3498c"
complement(5272..5324)
/gene="Rv3498c"
/note="Rv3498c. (MTV023.05c), len: 350. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.10
(346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic
region at N-terminus. FASTA scores: 297050|MTC128.10 (346
aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in
Query Match 8.9%; Score 40.2; DB 1; Length 47852;
Best Local Similarity 55.3%; Pred. No. 4.9;
Matches 78; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 203 AGCGTCACCTCCAGGGGGACCTGGCTACTTGCCTCCCGAGGCTTCCGCGTCCACGGGCTC 262
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31717 ACCGGTGCCTGTCGCGCGCGCCACCTTTTCGCGCGCGCGCGCTTACCGCGCGC 31658
QY 263 ACTTGTGGCTCCGCCCTGGCTGCTGGGATGTCGGCGCGCGAGACCATGTCACCTGCCAG 322
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31657 CCCCGGGTTCGCGCGCGTGCCTGATATTGCGCGCGCTTCCGCGCGTCCCGCGCGC 31598
QY 323 TCGCGGGGATGACGTGACCC 343
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31597 ACCCGCGGACCACTTGACC 31577

RESULT 8
HSE4 HSE4 5880 bp DNA VRL 03-FEB-1999
LOCUS Equine herpesvirus 4 (EHV4) genome, thymidine kinase (TK) and
DEFINITION glycoprotein H (gH) genes.
ACCESSION D14486 D00683 D00684
NID 9221820
VERSION D14486.1 GI:221820
KEYWORDS UL21 gene; UL24 gene; gH gene; glycoprotein; thymidine kinase.
SOURCE Equine herpesvirus 4 (strain 1942).
ORGANISM Equine herpesvirus 4
VIRUSES; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Varicellovirus.
REFERENCE 1 (sites)
AUTHORS Nicolson,L., Cullinane,A.A. and Onions,D.E.
TITLE The nucleotide sequence of an equine herpesvirus 4 gene homologue
of the herpes simplex virus 1 glycoprotein H gene
J. Gen. Virol. 71 (Pt 8), 1793-1800 (1990)
90362066
2 (sites)
REFERENCE Nicolson,L., Cullinane,A.A. and Onions,D.E.
TITLE The nucleotide sequence of the equine herpesvirus 4 thymidine
kinase gene
J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
90362067
3 (bases 1 to 5880)
REFERENCE Nicolson,L.
AUTHORS Unpublished (1993)
JOURNAL Submitted (10-SEP-1990) to DBJ by:
COMMENT Lesley Nicolson
Dept. Veterinary Pathology. University of Glasgow
Vet School
Bearsden Roda, Glasgow G61 1QH. Scotland
UK.
Phone: 041-339-8855
Fax: 041-330-5733.
Location/Qualifiers
FEATURES
```

```

source
1. .5880
/organism="Equine herpesvirus 4"
/db_xref="taxon:10331"
complement(99. .104)
/note="put. polyadenylation signal (UL24); putative"
polyA_signal
complement(115. .120)
/note="put. polyadenylation signal (UL24); putative"
polyA_signal
complement(116. .934)
/note="putative; ORF1 (UL24 homologue)"
CDS
/codon_start=1
/protein_id="BAA03377.1"
/db_xref="PID:d1003887"
/db_xref="PID:g221821"
/db_xref="GI:221821"
/translaton="MKRKQRLTARSRLRAGIRCHNRFFYNMVDLASAKKNVYGARL
ALPFLSELPATETKATMGVSLAFVNLGQRRPDCVCTVQFHGSDAGVCILIBKTC
RFSKNMTASKNLQKGMROLHDSCLLAKTLPFGSEILLAPLVFVAQGRMVLRL
VTRLSPQVYSNAAVLSCTISRLAEYSPPISESTRTRRCVTRRTNSRAFRKTTGSI
OPTQAKPAATAAVASLFSATAQANTTNAVGYQPATISLANPLAVNASLFAPK"
861. .865
/note="put. TATA box (TK); putative"
misc_feature
887
/note="put. transcription initiation site; putative"
CDS
949. .2007
/partial
/note="ORF2; (TK)"
/codon_start=1
/protein_id="BAA03378.1"
/db_xref="PID:d1003888"
/db_xref="PID:g221822"
/db_xref="GI:221822"
/translaton="MAACVPPGEAPRASGTPTRQVTVIVYLDGVYIGKSTGTVR
MAAASGSGTLPPEPMAYWTLFETDVISGYDTONRKOGLNVDAAALIAHYQ
MSRTTPLYLHDFTLPGNLSLORGTPDLTLVDFRHPVASTVCFPAARVLLGDSM
CALMAMVATLREPQGNIVVTNLNVEHRLRTRARIGQIDITLILATLRNYEML
VNTCHFURSRVNRDGGELPTSCGAYKRRATQMDAFQERVSPBELGDTLFLFKTEL
LDORGVILEVHAWALDMLKRLNLVFSADLSGTPROCAAVVESLLPLMSSTLSDE
SASALERAARTFNAEMGV"
2049. .2054
polyA_signal
/note="put. polyadenylation signal (TK); putative"
TATA_signal
2127. .2133
/note="putative"
TATA_signal
2197. .2202
/note="putative"
CDS
2225. .4792
/partial
/note="ORF3; (gH)"
/codon_start=1
/protein_id="BAA03379.1"
/db_xref="PID:d1003889"
/db_xref="PID:g221823"
/db_xref="GI:221823"
/translaton="MSQPYLKIALIVAAITVSAIPVMTPTVSTSPQOTKLUHYVNGT
WYHNNTFYTRIDRTEPYNVNNLSSTFEVAISERNFTVNTPLGASVFWILKSAL
NPKPHQPCIANPPEGPGRPCVNSTVSLFNDFNLEPMTKNLLEFEVLDPNTITGW
TFERSKVATKNGVGVLPSPPTSPDVNNTIRDDGTPKQHLSDIIDEHTTFVLDLON
TKLTYISPPAAVWPITAFHAGITVMGCDTQAIAYLNGPWGQISVNNPPELMIV
APNDVRARVNLPRRLEPPGPIYKAVYVLSGNTFYLGHSNKSISREVAAYPE
ESLDYRHLSLANDTLAMLAELSNGSKSDVSYLYRIARLAVATFSLAEVIRLSDY
MLQELADVDINLRLIPLVKNYAAGGTADSSYSSDVAMDQFEVAQRIKIVADIN
IENLRKPMYHRSLKSVAYSRKPLPNVAFANRLITAMYKAIDRTIWNSTMR
VLEFVGAAGASHVILDDGDLGLHAKHDSMFLSLNELLCTAMCTASHVYAGV
KLEVMAGLIAGGVQSLFVFSFCMASARFDLAEERVLDDLVSIPRLTDLNGL
EDDGTTHSYGRSAGNLSRIAYNFDAVRVTFPELASCSTKPLKVLVPLASNRSY
VITRTAENIGLTVSLDGNIAKPIVISYITNGCVSRATIRSYLDHPGHTQSCYVC
GSVMRYMASGAIMDLIYIDDKVELQLVAGENSTPAFNPKLTPSMNALLMPNGT
VTLMASFASYSAFKIPSTYLWASIGGLLALLIYIVIVKMLCGGVINNDYSLLENSE"
4777. .4782
/note="put. polyadenylation signal (gH); putative"
repeat_region
4893. .5315
/note="direct repeat region"
rep_origin
5395. .5403
/note="put. replication origin; putative"

```

```

rep_origin
5440. .5448
/note="put. replication origin; putative"
CDS
complement(5755. .5880)
/note="putative; ORF4 (UL21 homologue)"
/codon_start=1
/protein_id="BAA03380.1"
/db_xref="PID:d1003890"
/db_xref="PID:g221824"
/db_xref="GI:221824"
/translaton="NVSLLITAFDNLDRKGAAYLYLLSERLKRGGVPVHVNR"
BASE COUNT 1459 a 1377 c 1499 g 1545 t
ORIGIN

Query Match      8.8%; Score 40; DB 17; Length 5880;
Best Local Similarity 47.3%; Pred. No. 6.6;
Matches 121; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 180 GCAGCATGGCCCTGAGTCCAGAGCGCTCACCTCCAGGGGGGACCTGGCTACTTGCCTCC 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5004 GCGGCGGCGGCGGAGGCGTGGCGGGGCGGCGGAGGCGTCTGCGGCGGCGGCGGAGG 5063
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 240 GAGGCTTCCCGCTCACCGGCTGCACCTTGTGGCTCCGCTTGTGGCTCGTGGGATGTGCGCG 299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5064 GCTGTGCGGCGGCGGAGGCGTGTCTGCGGCGGCGGAGGCGTCTGCGGCGGCGGCGG 5123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 300 CCGAGACCATATGTCTACTCCAGTGGCGGGGATGACCTGGACCGGAGCGCGTGTCTGTC 359
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5124 AGGGCTGTGCGGCGGCGGAGGCGTGTCTGCGGCGGCGGAGGCGTGTCTGCGGCGGCG 5183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 360 GTGTGAGCCCTTGAAGTCCGCGCGACGCGGTGCACAGCGCGGCGGAGCGGCTCCAGGT 419
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5184 GCGAGGGCTGTCTGCGGCGGCGGAGGCGTGTCTGCTGCGGCGGCGGAGGCGTGTCTGCT 5243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 420 CCGGAGGGGTTGCGGG 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5244 GCGGCGGCGGCGGAGG 5259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
SFU08223 7082 bp DNA BCT 07-SEP-1994
LOCUS Streptomyces fradiae T59235 tylosin biosynthetic gene region,
DEFINITION cytochrome P-450, dTDP-glucose synthase, dTDP-glucose dehydratase,
and thioesterase genes, complete cds.
ACCESSION U08223
NID 9473596
VERSION U08223.1 GI:473596
KEYWORDS Streptomyces fradiae.
SOURCE Streptomyces fradiae.
ORGANISM Streptomyces fradiae.
REFERENCE 1 (bases 1 to 7082)
AUTHORS Merson-Davies, L.A. and Cundliffe, E.
TITLE Analysis of five tylosin biosynthetic genes from the tyLBA region
of the Streptomyces fradiae genome
JOURNAL Mol. Microbiol. 13, 349-355 (1994)
MEDLINE 95075319
REFERENCE 2 (bases 1 to 7082)
AUTHORS Merson-Davies, L.A.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-1994) Louise A. Merson-Davies, Biochemistry,
University of Leicester, Adrian Bldg., University Road, Leicester,
LE1 7RH, UK
FEATURES
Source
1. 7082
/organism="Streptomyces fradiae"
/strain="T59235"
/db_xref="taxon:1906"
944. .2197
/note="ORF1"
/codon_start=1
CDS

```

```
/transl_table=11
/product="cytochrome P-450"
/protein_id="AAA21341.1"
/db_xref="PID:g473597"
/db_xref="GI:473597"
/translation="MTTOLAEKPEPERTANSVHPLAOPAGARGILLEWFARARA
EAPVWDSRQAWQVRDDYLTSTNQLFSSDFSPVFPPEEAILMGLPFGFIDV
PPRHGLRKLKSOAFTPRATRIAEITRGLDGLREKQIDVSDLAYPLPIV
IAELGLPAERDRLFEWVDVILNNEGMEYNLPDPFSETMGPAIKWGDVLYLRRIAL
KREPTDMLSGSLIAEAEYRREGRRLTDEIVNIVALLTAGHISATLLGNLFLVLEHR
EAQALRADRLDIPCAIETLRYSPENNIFELKEDTDILGHKPMKAGOMVAVTASA
NRDSAFSDPDTDFVRRPNKMHSEGHGHIHCLGSFLARLEAKVLELFFDFSYRV
EHDEVEFEEDLFAIRLPVTVIRH"
2677..3843
/notes="ORF2, GTG start codon"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAA21342.1"
/db_xref="PID:g473598"
/db_xref="GI:473598"
/translation="MTGLPRPAVRVPFHDLRDVHAATGVSEIGGALLRVAARGYLL
GAELAAFEERAEVCGNAHCYAVGSLDDARLALWALGVGEGDEVIVSFHFIASWLA
VSAICATVPVPEGDPGPGFGLDDPDLRLEAALTPTRAVMPVHLVGHVPDLDPVG
AFAPGHLAVVDEDAQAARYKRRIGSGRTAFSPYKRNIGALGDGGVAVTSDPEL
ADRLRLNAGAREKYRHEERGTSRLDELQAAVLSVKLPLYLDAWNTRRREIATRYGE
ALAGLPVTVPEGRVAERVOYVLRSPYRDLRRLAEAGVETLVHVPVAVHAGSAY
AGAGPCPAGGLPRAERLAGEVLSLPIGPHLPDEAVEVIAVQSAALSWEQGP"
3840..4751
/notes="ORF3; GTG start codon"
/codon_start=1
/transl_table=11
/product="GTP-glucose synthase"
/protein_id="AAA21343.1"
/db_xref="PID:g473599"
/db_xref="GI:473599"
/translation="MNDPRRAMKGIILAGSGTRLRPLTGTLSKQLLPVYDKPMIY
PLSLMLAGIREIQIISCKHDLFRSLTGLEDGLGLSISYAEORPGRIAGFAEFLIG
RHIGDDALILGDNVFGPGFSSVLTGTVARLDGCELFYGVKDAHRVGEIDSG
RLLSLEPRPRLPCRLVLYTNDVVEIARTISPSARGELEITDVKVYLEQORAA
HGAGAVAWLMDGTHDSLLQAGQYVOLLERQGERIACIEIATMRMGFISAEQCYRUG
QELRSSYGSYIIDVAMGAAADSRAQ"
4812..5813
/notes="ORF4"
/codon_start=1
/transl_table=11
/product="GTP-glucose dehydratase"
/protein_id="AAA21344.1"
/db_xref="PID:g473600"
/db_xref="GI:473600"
/translation="MRVLVTGGAGFIGSHFTGQLLTGAYPDLGATRTVVLDKLTVAGN
PANLHVAGHPDLEVRGDIADHGWRLMEGVGLVHFAESHVDRSIESAEVFT
NVEGTRVLQAADVAGVGRFVHISTDEYVIAEGSWPEDHPVAPNSPYATKAASDL
LALAHRTYGLDVTRCSNNYPROYPEKAVPLFTNLLDGLPPLYGDGTEWL
HVDDCHGVALVAGGPGVINYIGGTELTNAELTDRLILECGADRSLARRVADRP
HDDRYSDVTTKIREELGYAPRTGITEGLAGTVAVYDRNRAWWEPLKRSPGGRELER"
5875..6636
/notes="ORF5, GTG start codon"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAA21345.1"
/db_xref="PID:g473601"
/db_xref="GI:473601"
/translation="MWLRKCYAPVGTPAHRLVCFPHAGGSARAYRPFPALEAAAGVET
HAVQYPGQRKRKEPTELEAERVLPELRLLDAPDGPVVALFGHSMGAVVAYET
ARLLHRSAGAPRAGLISGRAPTDRTCTAHLGDRELLAEIRLQGTDPICALADEE
VLRMLPAIRGDYAAVGYRHPVPGPRPCPLTVFTGTDADPNVTLPEAEAWRELITGAP
ALRVPGGHVLINDOREAVCTIETILRHGSKSAH"
1008 a 2479 c 2535 g 1060 t

BASE COUNT
ORIGIN

Query Match 8.8%; Score 39.8; DB 1; Length 7082;
Best Local Similarity 50.3%; Pred. No. 7.2;
Matches 98; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 196 GTGCCAGAGCGTCACCTCCAGGGGGACCTGCTACTTGTCCCGGAGGCTTCGCGGTAC 255
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3532 GTGCGGAGGCGCGCTGCGGAGCGCGTCTGGCATCAGTACGTGTGTCGACGCCGTAC 3591
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 256 CGGCTGCACCTGTGCTCGCCCTGTGGCTCGTGGGATGTGCGCGCGGAGACCATGTCA 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3592 CGGACCGGTTGCGGCGCGGCTGGCGGAGCGGGGTGGAGACCTGTGCTTATCCG 3651
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 CTGCCAGTCGCGGGGATGAGTGGACGCGCTGCTGTGTGTGCGAGCCCTGAGG 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3652 GTGGCGGTGCACGCGTGGGCGGTACGCGGCGCGGCGCGTGTCCCGCGGGGCTG 3711
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 TCGCGGCGAGCGGT 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3712 CGCGCGCGGAGCGT 3726
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
CVCMYC 1320 bp mRNA MAM 07-OCT-1996
LOCUS C.familiaris mRNA for c-myc proto-oncogene.
DEFINITION X95367
ACCESSION g1171522
NID X95367.1 GI:1171522
VERSION c-myc gene; DNA-binding protein.
KEYWORDS dog.
SOURCE Canis familiaris
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1320)
AUTHORS Pirson,I., Coulonval,K., Lamy,F. and Dumont,J.E.
TITLE c-Myc expression is controlled by the mitogenic CAMP-cascade in
thyrocytes
J. Cell. Physiol. 168 (1), 59-70 (1996)
MEDLINE 96257835
REFERENCE 2 (bases 1 to 1320)
AUTHORS Pirson,I.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1996) I. Pirson, IRIHBN-ULB, Blg C- (C4-123),
Campus Erasme, Route de Lennik 808, B- 1070 Brussels- BELGIUM,
BELGIUM
FEATURES
source
1..1320
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone_lib="lambda gt11"
/tissue_type="thyroid"
1..1320
/gene="c-myc"
1..1320
/gene="c-myc"
/notes="DNA-binding protein"
/codon_start=1
/protein_id="CAA64654.1"
/db_xref="PID:e220184"
/db_xref="PID:g1171523"
/db_xref="GI:1171523"
/db_xref="SWISS-PROT:Q28350"
/translation="MPLNVFSFANNRVDLDYDSVQPYFYCDEENFYQQQQQSELQPPA
PSEDLWKKFELLPTPLSPRSRGLSPSYAVAFSPRGGDDGGGSGFSTADQLEMY
TELLGDGMVQFCIDPDDETAKNIIDQCMWSGFSAAKLVSEKLSYQAARKDSG
SPSPARGGCGSTSLYLQDLSSAASECIDPSVVRPYPLNDSSSPKPSQSAATSP
SSDSLSSAESSPRASPELAHEETPTTSSDSEEOEDEIDVVSVEKKQAPAKR
SESGPSAGGSKSPHSPVLVLRCHVSTHOHNAPSPTRKDTPAAKRAKLDGRVLK
QLSNRPSKSSPSDTEENKRRTHNVLDERNRLRSFFALRDQIPLENNKAPK
VWLKATAYILSVQAEQKLSEKLDLKRREQLKHLQELRNSGA"
300 a 451 c 364 g 205 t

BASE COUNT
ORIGIN
```

```
Query Match      8.8%; Score 39.8; DB 3; Length 1320;
Best Local Similarity 48.1%; Pred. No. 8.4;
Matches 113; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 188 GGCCTGGAGTCCAGAGGCTCACCTCCAGGGGGACCTGGCTTACTTGGCCCCGAGGCTTC 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 GACATGGTGAACAGAGCTTCATCTCGACCGGAGCAGGAGACCTTCATCAAAACATC 384

QY 248 GCCGTACCGGCTGACATTTGGCTCCGCTGTCGCTGCTGGGATGTGGGCGCCGAGACC 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 ATCATCCAGGATGATGTGGAGCGCTTCGCGCGCGCCGCAAGCTCTCTCGGAGAG 444

QY 308 ACATGCTACTGCGAGTGGCGGCTGACTGACCGGAGCGGCTGCTGCTGTGTCAG 367
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 CTGGCTCTTACCAGGCTGCGCGAAGACAGCGACGCGCCGCGCTCGCGGGCCC 504

QY 368 CCCTGAGGTGCGCGCAGCGCTGCACAGCGGCGGAGGCGGCTCCAGGTCCG 422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 GCGGCTGCTCCACTCCAGCTGTACTGTGACGACCTGAGGCGCGCGCTCCG 559

RESULT 11
HVU49482      806 bp mRNA PLN 10-SEP-1996
LOCUS      Hordeum vulgare low temperature-responsive RNA-binding protein
DEFINITION      (bit801) mRNA, complete cds.
ACCESSION      U49482
NID      g1229137
VERSION      U49482.1 GI:1229137
KEYWORDS      barley.
SOURCE      Hordeum vulgare
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Hordeum.
REFERENCE      1 (bases 1 to 806)
AUTHORS      Dunn,M.A., Brown,K., Lightowlers,R. and Hughes,M.A.
TITLE      A low-temperature-responsive gene from barley encodes a protein
with single-stranded nucleic acid-binding activity which is
phosphorylated in vitro
JOURNAL      Plant Mol. Biol. 30 (5), 947-959 (1996)
MEDLINE      96270373
REFERENCE      2 (bases 1 to 806)
AUTHORS      Dunn,M.A. and Hughes,M.A.
TITLE      Direct Submission
JOURNAL      Submitted (20-FEB-1996) M. Allison Dunn, Biochemistry and Genetics,
University of Newcastle upon Tyne, Medical School, Framlington
Place, Newcastle upon Tyne NE2 4HH, UK
FEATURES      Location/Qualifiers
               1..806
               /organism="Hordeum vulgare"
               /cultivar="Igr1"
               /db_xref="taxon:4513"
               /tissue_type="leaf meristem"
               70..555
               /gene="b1t801"
               /note="glycine-rich RNA-binding protein"
               /codon_start=1
               /product="low temperature-responsive RNA-binding protein"
               /protein_id="AA07749.1"
               /db_xref="PID:g1229138"
               /db_xref="GI:1229138"
               /translation="MADVEYRCFVGLRWATDDSLONAFSKYGDVIDSKIITDRETG
               ESRGFGVTFASDEAMQAIEMNGQDLDRNITWNEAQSRSDGGGFGGGGGYGG
               QRRREGGGYGGGGGGTGGGRSGGGGTGSRDGGGGYGGGGGGTGGSRGGGGGNWR
               E"
BASE COUNT      121 a 243 c 286 g 156 t
ORIGIN
```



```

repeat_region complement(10863..10925)
repeat_region /rpt_family="GC-rich"
repeat_region 11162..11208
repeat_region /rpt_family="LINE2"
repeat_region complement(11339..11368)
repeat_region /rpt_family="GARA)n"
repeat_region complement(11369..11666)
repeat_region /rpt_family="AluSx"
repeat_region complement(11972..12274)
repeat_region /rpt_family="AluSq"
repeat_region 12696..12797
repeat_region /rpt_family="MIR"
repeat_region 13059..13138
repeat_region /rpt_family="MIR"
repeat_region 13220..13331
repeat_region /rpt_family="LINE2"
repeat_region complement(13416..13723)
repeat_region /rpt_family="AluY"
repeat_region complement(13730..14018)
repeat_region /rpt_family="AluJo"
repeat_region complement(14074..14250)
repeat_region /rpt_family="AluJo"
repeat_region complement(14254..14551)
repeat_region /rpt_family="AluSq"
repeat_region complement(14566..14867)
repeat_region /rpt_family="AluY"
repeat_region complement(14888..15027)
repeat_region /rpt_family="FRAM"
repeat_region complement(15036..15149)
repeat_region /rpt_family="MIR"
repeat_region 15988..16261
repeat_region /rpt_family="AluJo"
repeat_region 16276..16398
repeat_region /rpt_family="AluSq/x"
repeat_region complement(16583..16879)
repeat_region /rpt_family="AluSq"
repeat_region complement(17259..17285)
repeat_region /rpt_family="(GA)n"
repeat_region complement(17286..17598)
repeat_region /rpt_family="AluJo"
repeat_region complement(17603..18207)
repeat_region /rpt_family="Li"
repeat_region complement(18217..18348)
repeat_region /rpt_family="FLAM_C"
repeat_region complement(18357..18424)
repeat_region /rpt_family="LINE2"
repeat_region complement(19079..19234)
repeat_region /rpt_family="FRAM"
repeat_region complement(19597..19659)
repeat_region /rpt_family="(GGGA)n"
repeat_region complement(19661..19962)
repeat_region /rpt_family="AluSx"
repeat_region 20116..20202
repeat_region /rpt_family="MIR"
repeat_region 20336..20550
repeat_region /note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: good, score: 51.000"
repeat_region complement(20724..20837)
repeat_region /rpt_family="(GA)n"
repeat_region 20895..21067
repeat_region /rpt_family="AluJo"
repeat_region 21343..21639
repeat_region /rpt_family="AluSq"
repeat_region complement(21725..21831)
repeat_region /rpt_family="MIR"
repeat_region 23022..23156
repeat_region /rpt_family="AluSq/x"
repeat_region 23157..23443
repeat_region /rpt_family="AluSx"
repeat_region complement(23448..23576)
repeat_region /rpt_family="FLAM_C"
repeat_region complement(24313..24455)
repeat_region /note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 93.000--DDS
similarity to overlapping ESTs:~(24455..24153) AA121509
zk88c10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA
clone 489906 3' similar to WP: ZK593.7 CE06628 YEAST
JTA107 LIKE ; (303..1); 99% identity.--(24455..24154)
AA215299 zF94B05.F1 NCI_CGAP_GCB1 Homo sapiens cDNA clone
IMAGE:683313 5' similar to WP:ZK593.7 CE06628 YEAST JTA107
LIKE ; (165..1); 88% identity.--Additional EST
matches:-A1018313, A1004446, AA989671, A1036173"
complement(join(24313..24455,26758..26825,31020..31095,
31194..31344))
CDS
/note="Hypothetical human protein most similar to WP:
ZK593.7 CE06628"
/codon_start=1
/evidence=not_experimental
/product="R30783.1"
/protein_id="AAC25622.1"
/db_xref="PID:g3289993"
/db_xref="GI:3289993"
/translation="MSAGARRNORFSPNAPRATHAPRMVGEKGVAPAKSH
TARODGEGEHLGLVQVHRDDPGKVPGRPSGILKGFDPDLNLVDGIEYMRDPDD
QYKLTEDPROLGLVVCRTSVLIPCQDGEAIPNPFIOQDA"
24891..25193
repeat_region /rpt_family="AluY"
repeat_region complement(25347..25649)
repeat_region /rpt_family="AluJb"
repeat_region 25974..26083
repeat_region /rpt_family="AluJo/FLAM"
repeat_region complement(26087..26384)
repeat_region /rpt_family="AluSx"
repeat_region complement(26397..26433)
repeat_region /rpt_family="AT-rich"
repeat_region complement(26708..26825)
misc_feature
Query Match 8.7%; Score 39.2; DB 11; Length 42189;
Best Local Similarity 47.9%; Pred. No. 8.2;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 190 CTTGGAGTCCAGAGCGTCACCTCCAGGGGGGACCTGGCTACTTGGCCCGAGGCTTCGC 249
Db 10975 CCGCGGCTCCAGTGTCCCGGCACCTGCGCCACGTGGCTCTCACACCTGCGGCCCG 10916
QY 250 CGTCAACCGGCTGCACCTGTGGCTCCCGCTGTGGGTGTGCGCGCCGAGACAC 309
Db 10915 CGCGCGGACGCGGCGCCAGCGAGCCCGCGCCCGCCGCGCGCGCGCGAGGAC 10856
QY 310 ATGTCACTGCCAGTGCAGCGGCGATGACCTGGACCGAGCGCGCTGTGTGTCGACGC 369
Db 10855 CGCGCGGCGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10796
QY 370 CTGAGGTCCGCGCAGCGCGTGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 425
Db 10795 GCAGCGCGCGCGCGCGCGCGCGCTGCTGCTGCGGCTCGCGCGCGCGCGCGCGCG 10740
Search completed: October 1, 1999, 15:30:36
Job time: 12752 sec

```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 10:59:16 ; Search time 14.12 Seconds
(without alignments)
181.169 Million cell updates/sec

Title: US-09-099-898-2
Perfect score: 581
Sequence: 1 MKALCILLLPVLGLVSSKT.....CHCOCAGMDWTGACRCRVQ 108
Scoring table: BLOSUM62

Searched: 188963-seqs, 23686106 residues

Database: A_Geneseq 36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	581	100.0	108	1 W87710	A cysteine rich so
2	581	100.0	109	1 Y12933	Amino acid sequenc
3	318	54.7	114	1 W87707	A cysteine rich so
4	315	54.2	114	1 W87708	A cysteine rich so
5	284.5	49.0	111	1 W87709	A cysteine rich so
6	273.5	47.1	105	1 W87706	A cysteine rich so
7	245.5	42.3	111	1 W87705	A cysteine rich so
8	227	39.1	52	1 Y12613	Human 5' EST seque
9	202.5	34.9	111	1 W87704	A cysteine rich so
10	82	14.1	2321	1 W49698	Human Notch3 prote
11	82	14.1	1872	1 W68510	Partial human Notc
12	76.5	13.2	60	1 R98207	Nucleotide used in
13	73	12.6	131	1 W4444	Mouse Tie receptor
14	72	12.4	1193	1 R91427	Kalinin/laminin 5
15	72	12.4	1111	1 R91428	Kalinin/laminin 5
16	71.5	12.3	519	1 W30826	The novel tyrosina
17	71.5	12.3	2476	1 W67738	Pig p105 zona pell
18	71	12.2	397	1 R27139	ALP of Lysobacter
19	69	11.9	1094	1 R39821	Truncated tie rece
20	68.5	11.8	2471	1 Y06816	Human Notch2 (humN
21	68	11.7	1138	1 R39820	tie receptor kinas
22	68	11.7	1122	1 R73954	Human tie tyrosine
23	68	11.7	102	1 W24566	Serine protease C-
24	67	11.5	1801	1 W50895	Rat laminin B2 cha
25	67	11.5	915	1 Y13350	Amino acid sequenc
26	66.5	11.4	488	1 R67757	Human fetal brain
27	66.5	11.4	524	1 R67758	Human fetal brain
28	66.5	11.4	670	1 R67759	Human fetal brain
29	66.5	11.4	769	1 R75322	Human fetal brain
30	66	11.4	177	1 R40167	Recombinant growth
31	66	11.4	297	1 W20066	Thielavia terrestr
32	65.5	11.3	1080	1 P50296	Mouse epidermal gr
33	65.5	11.3	1964	1 W95557	Mus musculus notch
34	65	11.2	289	1 R52633	Guinea pig PH-30,
35	65	11.2	2707	1 W27161	Mouse receptor ME2
36	65	11.2	1148	1 W87895	Human JAGGED2 prot
37	64.5	11.1	375	1 W44141	Thermotable alkali
38	64.5	11.1	2799	1 W81867	Human tumour suppr
39	64	11.0	688	1 R80361	Tick WGL+ antigen.
40	64	11.0	213	1 R45359	Wheat germ aggluti
41	64	11.0	197	1 W36951	Protein encoded by
42	64	11.0	685	1 W80813	Nucleotide sequenc
43	64	11.0	659	1 W94497	Human delta-2 matu

44 64 11.0 685 1 W94507 Human delta-2 prot
45 64 11.0 500 1 W94496 Human delta-2 prot

ALIGNMENTS

RESULT 1

W87710 ID W87710 standard; Protein; 108 AA.

AC W87710;

DI 09-WAR-1999 (first entry)

DE A cysteine rich soluble protein designated C23.

KW Cysteine rich soluble protein; CRSP; C23; cell development;

KW mammalian immune system; antibody; abnormal proliferation; cancer;

OS inflammation; degeneration; regeneration; degeneration; atrophy.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Protein

FT 1..18 /note= "mature protein"

PN W09858061-A1.

PD 23-DEC-1998.

PR 18-JUN-1998; U12236.

PR 09-OCT-1997; US-061641.

PR 19-JUN-1997; US-878730.

PR 19-JUN-1997; US-878878.

PA (SCHE) SCHERING CORP.

PI Franz-Bacon K, Gorman DM, McClanahan TK;

DR WPI: 99-095339/08.

DR N-PSDB: V84059.

PT New cysteine-rich soluble proteins - used to modulate proliferation, differentiation, trafficking and development of cells, e.g. for treating inflammation, cancer and degeneration

PS Claim 1: Page 18-19; 119pp: English.

CC The present sequence represents a cysteine rich soluble protein (CRSP) designated C23.

CC modulate Physiology, differentiation, trafficking and development of cells (including those in culture), particularly cells of the mammalian immune system. They are used for treatment of abnormal proliferation (cancer, inflammation or degeneration), regeneration, degeneration and atrophy. The proteins are also used to raise, or detect, antibodies, to design oligonucleotides for library screening, in drug screens and to isolate cognate receptors. The antibodies are used for affinity purification of CRSP, to screen expression libraries, to identify CRSP-expressing cells, as diagnostic immunoassay reagents, to produce anti-idiotypic antibodies (useful for diagnosis), in competitive drug screens, and as therapeutic modulators.

CC Sequence 108 AA;

Query Match 100.0%; Score 581; DB 1: Length 108;
Best Local Similarity 100.0%; Pred. No. 8.3e-50; Mismatches 0; Indels 0; Gaps 0;
Matches 108; Conservative 0;

QY 1 MKALCILLLPVLGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIGLCQSVTSRGDL 60

Db 1 MKALCILLLPVLGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIGLCQSVTSRGDL 60

QY 61 ATCPRGFAVTGCTCGSACGSDVRAETTCCHCCAGMDWTGACRCRVQ 108

Db 61 ATCPRGFAVTGCTCGSACGSDVRAETTCCHCCAGMDWTGACRCRVQ 108

RESULT 2

Y12933

ID Y12933 standard; Protein; 109 AA.

AC Y12933;

DI 17-JUN-1999 (first entry)

DE Amino acid sequence of a human secreted peptide.

KW Human secreted protein; cancer; immune disorder; infection;

KW inflammatory disorder; skin disorder; tumour; atherosclerosis;

KW restenosis; autoimmunity; Alzheimer's disease;

KW peripheral neuropathy; trauma; spinal cord injury; allergy;

same as protein

ABW

App 12/2/95

ABW

KW • hematopoietic disorder; skeletal disorder; neurological disorder;
KW arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW transplant rejection; ss.

OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..23
FT Protein /note= "signal peptide"
FT 24..108
FT /note= "secreted protein"

PN WO9911293-A1.
PD 11-MAR-1999.
PF 03-SEP-1998; U18360.
PR 12-SEP-1997; US-058974.
PR 05-SEP-1997; US-057626.
PR 05-SEP-1997; US-057663.
PR 05-SEP-1997; US-057669.
PR 12-SEP-1997; US-058666.
PR 12-SEP-1997; US-058667.
PR 12-SEP-1997; US-058973.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
PI Rosen GA, Ruben SM, Shi Y;
DR WPI: 99-204988/17.
DR N-PSDB: X51720.
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. neurological disorders,
PT tumours, immune disorders, inflammation or haematological disorders
PS Claim 11; Page 190; 215pp; English.
CC Y12914-68 represent human secreted proteins. The polypeptides and
CC their corresponding polynucleotides are useful for preventing,
CC treating or ameliorating medical conditions, e.g. by protein or gene
CC therapy. Pathological conditions can also be diagnosed by determining
CC the amount of the new polypeptides in a sample or by determining the
CC presence of mutations in the new polynucleotides. Specific uses are
CC described for each polynucleotide, based on which tissues they are
CC most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, immune disorders, infection,
CC inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
CC disorders, skeletal disorders, neurological disorders, arthritic
CC disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC rejection. The polypeptides are also useful for identifying their
CC binding partners.
SQ Sequence 109 AA;

Query Match 100.0%; Score 581; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 8.4e-50;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKALCLLLPVLGLVSSKTLCSMEAINERTQEVAGSLIFRAISSIGLEQCQSVTSRGDL 60
DB 1 MKALCLLLPVLGLVSSKTLCSMEAINERTQEVAGSLIFRAISSIGLEQCQSVTSRGDL 60
QY 61 ATCPRGFVTCGTCGACGSDVRAETTCCHCCQAGMDWTGACRCVQ 108
DB 61 ATCPRGFVTCGTCGACGSDVRAETTCCHCCQAGMDWTGACRCVQ 108

RESULT 3
W87707
ID W87707 standard; Protein; 114 AA.
AC W87707;
DT 09-MAR-1999 (first entry)
DE A cysteine rich soluble protein designated C19.
KW Cysteine rich soluble protein; CRSP; C19; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy.
OS Mus sp.
FH Key Location/Qualifiers
FT Protein 1..20
FT /note= "mature protein"

PN WO9858061-A1.
PD 23-DEC-1998.
PF 18-JUN-1998; U12236.
PR 09-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PA (SCHE) SCHERING CORP.
PI Franz-Bacon K, Gorman DM, McManahan TK;
DR WPI: 99-095339/08.
DR N-PSDB: V84056.
PT New cysteine-rich soluble proteins - used to modulate proliferation,
PT differentiation, trafficking and development of cells, e.g. for
PT treating inflammation, cancer and degeneration
PS Claim 1; Page 15; 119pp; English.
CC The present sequence represents a cysteine rich soluble protein (CRSP)
CC designated C19. CRSP proteins, and their (ant)agonists, are used to
CC modulate physiology, differentiation, trafficking and development
CC of cells (including those in culture), particularly cells of the
CC mammalian immune system. They are used for treatment of abnormal
CC proliferation (cancer, inflammation or degeneration), regeneration,
CC degeneration and atrophy. The proteins are also used to raise or
CC detect antibodies, to design oligonucleotides for library screening,
CC in drug screens and to isolate cognate receptors. The antibodies are
CC used for affinity purification of CRSP, to screen expression libraries,
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC to produce anti-idiotypic antibodies (useful for diagnosis), in
CC competitive drug screens, and as therapeutic modulators.
SQ Sequence 114 AA;

Query Match 54.7%; Score 318; DB 1; Length 114;
Best Local Similarity 55.8%; Pred. No. 3.4e-24;
Matches 58; Conservative 16; Mismatches 30; Indels 0; Gaps 0;
QY 4 LCLLLPVLGLVSSKTLCSMEAINERTQEVAGSLIFRAISSIGLEQCQSVTSRGDLATC 63
DB 8 LLFLFLPELLGSSMPLCPIDDAIDKKIKQDFNLPNAIKNLGNLWTVSSRGKLASC 67
QY 64 PRGFVTCGTCGACGSDVRAETTCCHCCQAGMDWTGACRCVQ 107
DB 68 PEGTAVLSCGACGSDVREKVCQCQARIDWTAARCKLQ 111

RESULT 4
W87708
ID W87708 standard; Protein; 114 AA.
AC W87708;
DT 09-MAR-1999 (first entry)
DE A cysteine rich soluble protein designated C19.
KW Cysteine rich soluble protein; CRSP; C19; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy.
OS Rattus sp.
FH Key Location/Qualifiers
FT Protein 1..17
FT /note= "mature protein"

PN WO9858061-A1.
PD 23-DEC-1998.
PF 18-JUN-1998; U12236.
PR 09-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PA (SCHE) SCHERING CORP.
PI Franz-Bacon K, Gorman DM, McManahan TK;
DR WPI: 99-095339/08.
DR N-PSDB: V84057.
PT New cysteine-rich soluble proteins - used to modulate proliferation,
PT differentiation, trafficking and development of cells, e.g. for
PT treating inflammation, cancer and degeneration
PS Claim 1; Page 16; 119pp; English.
CC The present sequence represents a cysteine rich soluble protein (CRSP)
CC designated C19. CRSP proteins, and their (ant)agonists, are used to
CC modulate physiology, differentiation, trafficking and development

CC of cells (including those in culture), particularly cells of the
 CC mammalian immune system. They are used for treatment of abnormal
 CC proliferation (cancer, inflammation or degeneration), regeneration,
 CC degeneration and atrophy. The proteins are also used to raise, or
 CC detect, antibodies, to design oligonucleotides for library screening,
 CC in drug screens and to isolate cognate receptors. The antibodies are
 CC used for affinity purification of CRSP, to screen expression libraries,
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.
 SQ Sequence 114 AA;

Query Match 54.2%; Score 315; DB 1; Length 114;
 Best Local Similarity 54.1%; Pred. No. 6.6e-24;
 Matches 60; Conservative 16; Mismatches 31; Indels 4; Gaps 1;

QY 1 MKALCLLLL---PVLGGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIGLEQSVTS 56
 DB 1 MKNLSFLFLFLVGLGFGSMSCLPDDEALSKINQDFSSLLPAAMKNTVLRHWSVSS 60

QY 57 RGLATCPRGFAVTCGTCGACGSDVRAETTCCHQCAGMDWTGARCCRVQ 107
 DB 61 RGRLASCEGTIVTSCGCGSGSDVREDTMCCHQCQCSIDWTARCCCLR 111

RESULT 5
 W87709
 ID W87709 standard; Protein; 111 AA.
 AC W87709;
 DT 09-MAR-1999 (first entry)
 DE A cysteine rich soluble protein designated C10.
 KW Cysteine rich soluble protein; CRSP; C10; cell development;
 KW mammalian immune system; antibody; abnormal proliferation; cancer;
 KW inflammation; degeneration; regeneration; degeneration; atrophy.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Protein 1..20
 FT /note= "mature protein"
 PN W09858061-A1.
 PD 23-DEC-1998.
 PF 18-JUN-1998; U12236.
 PR 09-OCT-1997; US-061641.
 PR 19-JUN-1997; US-878730.
 PR 19-JUN-1997; US-878878.
 PA (SCHE) SCHERING CORP.
 PI Franz-Bacon K, Gorman DM, McClanahan TK;
 DR WPI; 99-095339/08.
 DR N-PSDB; V84058.
 PT New cysteine-rich soluble proteins - used to modulate proliferation,
 PT differentiation, trafficking and development of cells, e.g. for
 PT treating inflammation, cancer and degeneration
 PS Claim 1; Page 17; 119pp; English.
 CC The present sequence represents a cysteine rich soluble protein (CRSP)
 CC designated C10. CRSP proteins, and their (ant)agonists, are used to
 CC modulate physiology, differentiation, trafficking and development
 CC of cells (including those in culture), particularly cells of the
 CC mammalian immune system. They are used for treatment of abnormal
 CC proliferation (cancer, inflammation or degeneration), regeneration,
 CC degeneration and atrophy. The proteins are also used to raise, or
 CC detect, antibodies, to design oligonucleotides for library screening,
 CC in drug screens and to isolate cognate receptors. The antibodies are
 CC used for affinity purification of CRSP, to screen expression libraries,
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.
 SQ Sequence 111 AA;

Query Match 49.0%; Score 284.5; DB 1; Length 111;
 Best Local Similarity 49.0%; Pred. No. 5.9e-21;
 Matches 51; Conservative 21; Mismatches 31; Indels 1; Gaps 1;

QY 4 ICLLLPVLGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIGLEQSVTSRGDLAT 62
 DB 7 LLLIILPLQLINPGSTQCSSLSDVMDKKIKDLVLSLEYSPSPISKKLSCASVKSQGRPS 66

QY 63 CPRGFAVTCGTCGACGSDVRAETTCCHQCAGMDWTGARCCRV 106
 DB 67 CPAGNAVTCACGYCGGSDVQLETTCHQCQCSVDWWTARCCHL 110

RESULT 6
 W87706
 ID W87706 standard; Protein; 105 AA.
 AC W87706;
 DT 09-MAR-1999 (first entry)
 DE A cysteine rich soluble protein designated C18.
 KW Cysteine rich soluble protein; CRSP; C18; cell development;
 KW mammalian immune system; antibody; abnormal proliferation; cancer;
 KW inflammation; degeneration; regeneration; degeneration; atrophy.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT Protein 1..19
 FT /note= "mature protein"
 PN W09858061-A1.
 PD 23-DEC-1998.
 PF 18-JUN-1998; U12236.
 PR 09-OCT-1997; US-061641.
 PR 19-JUN-1997; US-878730.
 PR 19-JUN-1997; US-878878.
 PA (SCHE) SCHERING CORP.
 PI Franz-Bacon K, Gorman DM, McClanahan TK;
 DR WPI; 99-095339/08.
 DR N-PSDB; V84055.
 PT New cysteine-rich soluble proteins - used to modulate proliferation,
 PT differentiation, trafficking and development of cells, e.g. for
 PT treating inflammation, cancer and degeneration
 PS Claim 1; Page 14; 119pp; English.
 CC The present sequence represents a cysteine rich soluble protein (CRSP)
 CC designated C18. CRSP proteins, and their (ant)agonists, are used to
 CC modulate physiology, differentiation, trafficking and development
 CC of cells (including those in culture), particularly cells of the
 CC mammalian immune system. They are used for treatment of abnormal
 CC proliferation (cancer, inflammation or degeneration), regeneration,
 CC degeneration and atrophy. The proteins are also used to raise, or
 CC detect, antibodies, to design oligonucleotides for library screening,
 CC in drug screens and to isolate cognate receptors. The antibodies are
 CC used for affinity purification of CRSP, to screen expression libraries,
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.
 SQ Sequence 105 AA;

Query Match 47.1%; Score 273.5; DB 1; Length 105;
 Best Local Similarity 48.6%; Pred. No. 6.5e-20;
 Matches 51; Conservative 12; Mismatches 35; Indels 7; Gaps 2;

QY 4 ICLLLPVLGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIGLEQSVTSRGDLA 61
 DB 5 LCLFLVSLFLPLVPGNACSCFSLVDQIRKALSRQPEKTTIS-----CTSSTSGRLA 59

QY 62 TCPRGFAVTCGTCGACGSDVRAETTCCHQCAGMDWTGARCCRV 106
 DB 60 SCAPGVVTCGACGYCGGSDVIRNGNTCHQCQCSVMDWASARCCRM 104

RESULT 7
 W87705
 ID W87705 standard; Protein; 111 AA.
 AC W87705;
 DT 09-MAR-1999 (first entry)
 DE A cysteine rich soluble protein designated C2b.
 KW Cysteine rich soluble protein; CRSP; C2b; cell development;
 KW mammalian immune system; antibody; abnormal proliferation; cancer;

KW inflammation; degeneration; regeneration; degeneration; atrophy.

OS Mus sp.

FH Key

FT Protein

PN W09858061-A1.

PD 23-DEC-1998.

PF 18-JUN-1998; U12236.

PR 09-OCT-1997; US-061641.

PR 19-JUN-1997; US-878730.

PR 19-JUN-1997; US-878878.

PA (SCHE) SCHERING CORP.

PI Franz-Bacon K, Gorman DM, McClanahan TK;

DR WPI; 99-095339/08.

DR N-PSDB; V84053.

PT New cysteine-rich soluble proteins - used to modulate proliferation,

PT differentiation, trafficking and development of cells, e.g. for

PT treating inflammation, cancer and degeneration

PS Claim 1; Page 13; 119pp; English.

CC The present sequence represents a cysteine rich soluble protein (CRSP)

CC designated C2b. CRSP proteins, and their (antagonists, are used to

CC modulate physiology, differentiation, trafficking and development

CC of cells (including those in culture), particularly cells of the

CC mammalian immune system. They are used for treatment of abnormal

CC proliferation (cancer, inflammation or degeneration), regeneration,

CC detect, antibodies, to design oligonucleotides for library screening,

CC in drug screens and to isolate cognate receptors. The antibodies are

CC used for affinity purification of CRSP, to screen expression libraries,

CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,

CC to produce anti-idiotypic antibodies (useful for diagnosis), in

CC competitive drug screens, and as therapeutic modulators.

SQ Sequence 111 AA;

Query Match 42.3%; Score 245.5; DB 1; Length 111;

Best Local Similarity 40.2%; Pred. No. 3.7e-17;

Matches 43; Conservative 21; Mismatches 33; Indels 11; Gaps 2;

Qy 6 LLLLPVGLLVSSKTLCSMEAINRIQEVAG-----SLIFRAISSIGLEQCVTSRGD 59

Db 9 L I C I S L L Q L M P V N T E G L S T I E V K K E L L A N R D C P S T V T K T F S -----C T S I T A S G R 63

Qy 60 L A T C P R G F A V T G C T G S A G S G S D V A E T T C H C Q A G M D W T G A R C C R V 106

Db 64 L A S C P S G M T V T G C A G Y G G S W D I R D G N T C H C Q C S T M D M A T A R C C Q L 110

RESULT 8

Y12613

ID Y12613 standard; Protein; 52 AA.

AC Y12613;

DT 22-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO: 278 from WO 9906553.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; antiinflammatory; tumour inhibition; antitumour.

OS Homo sapiens.

PN W09906553-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; IB1237.

PR 01-AUG-1997; US-905051.

PA (GEST) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI; 99-153783/13.

DR N-PSDB; X41471.

PT New nucleic acids encoding human secreted proteins - obtained from

PT cDNA libraries derived from umbilical cord, lymph ganglia,

PT lymphocytes and placental tissue

PS Claim 34; Page 375; 411pp; English.

Qy

CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human

CC secreted proteins, and encode the proteins given in Y12521 to Y12668,

CC respectively. The proteins given represent the signal peptide and an

CC N-terminal fragment of a secreted protein. The nucleic acid sequences

CC can be used for producing secreted human gene products. They can also

CC be used to develop products for diagnosis and therapy. The proteins

CC obtained may have cytokine activity, cell proliferation/differentiation

CC activity, haematopoiesis regulating activity, tissue growth regulating

CC activity, reproductive hormone regulating activity, chemotactic/

CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/

CC ligand activity, antiinflammatory activity, tumour inhibition activity

CC or other activities. The products can be used in forensic, gene therapy

CC and chromosome mapping procedures. The sequences can also be used for

CC obtaining corresponding promoter sequences. The nucleic acids encoding

CC the signal peptide can be used for directing extracellular secretion of

CC a polypeptide or the insertion of a polypeptide into a membrane, or

CC importing a polypeptide into a cell.

SQ Sequence 52 AA;

Query Match 39.1%; Score 227; DB 1; Length 52;

Best Local Similarity 100.0%; Pred. No. 1e-15;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKALCILLLPVGLLVSSKTLCSMEAINRIQEVAGSLIFRAISSIG 48

Db 1 MKALCILLLPVGLLVSSKTLCSMEAINRIQEVAGSLIFRAISSIG 48

RESULT 9

W87704

ID W87704 standard; Protein; 111 AA.

AC W87704;

DT 09-MAR-1999 (first entry)

DE A cysteine rich soluble protein designated C2.

KW Cysteine rich soluble protein; CRSP; C2; cell development;

KW mammalian immune system; antibody; abnormal proliferation; cancer;

KW inflammation; degeneration; regeneration; degeneration; atrophy.

OS Mus sp.

FH Key

FT Protein

PN W09858061-A1.

PD 23-DEC-1998.

PF 18-JUN-1998; U12236.

PR 09-OCT-1997; US-061641.

PR 19-JUN-1997; US-878730.

PR 19-JUN-1997; US-878878.

PA (SCHE) SCHERING CORP.

PI Franz-Bacon K, Gorman DM, McClanahan TK;

DR WPI; 99-095339/08.

DR N-PSDB; V84053.

PT New cysteine-rich soluble proteins - used to modulate proliferation,

PT differentiation, trafficking and development of cells, e.g. for

PT treating inflammation, cancer and degeneration

PS Claim 1; Page 12-13; 119pp; English.

CC The present sequence represents a cysteine rich soluble protein (CRSP)

CC designated C2. CRSP proteins, and their (antagonists, are used to

CC modulate physiology, differentiation, trafficking and development

CC of cells (including those in culture), particularly cells of the

CC mammalian immune system. They are used for treatment of abnormal

CC proliferation (cancer, inflammation or degeneration), regeneration,

CC degeneration and atrophy. The proteins are also used to raise, or

CC detect, antibodies, to design oligonucleotides for library screening,

CC in drug screens and to isolate cognate receptors. The antibodies are

CC used for affinity purification of CRSP, to screen expression libraries,

CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,

CC to produce anti-idiotypic antibodies (useful for diagnosis), in

CC competitive drug screens, and as therapeutic modulators.

SQ Sequence 111 AA;

Query Match 34.9%; Score 202.5; DB 1; Length 111;

```
Best Local Similarity 36.3%; Pred. No. 5,6e-13;
Matches 37; Conservative 24; Mismatches 40; Indels 1; Gaps 1;

QY 6 LLLPLVGLLVSSKTLCSMEEAINEIRIOE-VAGSLIFRAISIGLEQSVTSRGDLATCP 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LICISLLQLWPVNTDETIEIIVENKVELLANPANYSTVTKTLSCITSVKTMNRWASCP 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 55 RGFVATGCTCSAGSGSDVRAETTCCHQCAGMDWTGARCRRV 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 AGMTATGCACGFACGSGWEIQSGDTCNCLLLVDWTARCCOL 110.

RESULT 10
W49698
ID W49698 standard; Protein; 2321 AA.
AC W49698;
DT 21-DEC-1998 (first entry)
DE Human Notch3 protein.
KW Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leukoencephalopathy; therapy.
OS Homo sapiens.
PN FR2751986-A1.
PD 06-FEB-1998.
PF 16-APR-1997; 004680.
PR 01-AUG-1996; FR-009733.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserre E;
DR WPI; 98-133138/13.
DR N-PSDB; V57001.
PT Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leukoencephalopathy
PS Claim 2; Fig 1.1-1.8; 45pp; French.
CC This sequence represents the human Notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of
CC the cerebral autosomal dominant arteriopathy with subcortical infarcts
CC and leukoencephalopathy (CADASIL) type. Blocking expression of a
CC mutated Notch3 gene or by substitution therapy with non-mutated Notch3
CC gene or protein can be used to treat CADASIL or related disorders.
SQ Sequence 2321 AA;

Query Match 14.1%; Score 82; DB 1; Length 2321;
Best Local Similarity 29.9%; Pred. No. 8.4;
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps 4;

QY 47 IGLEQSVTSRGDLATCPRGFAVTGCTCGSAGSGWDVRAETTC----- 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1300 VGVPCQQ-TPRGPRCACPPG--LSGPFSCRSPGPGSNASCAAPCLHGSGCRPAPLA 1356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 ---HCOACAGMDWTGARC 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1357 PFFRCACA-QGWTGPRC 1372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
W68510
ID W68510 standard; Protein; 1872 AA.
AC W68510;
DT 06-JAN-1999 (first entry)
DE Partial human Notch-3 protein.
KW Human; Notch3; transmembrane receptor; lateral inhibition; regulation;
KW developmental cascade; neurogenic gene; mutant; neurological disorder;
KW cerebral autosomal dominant arteriopathy; subcortical infarct; CADASIL;
KW leukoencephalopathy.
OS Homo sapiens.
PN Location/Qualifiers
FT Key
FT Misc_difference 328
FT /note= "encoded by NAN"

FT Misc_difference 401 /note= "encoded by GNN"
FT Misc_difference 403 /note= "encoded by GNC"
FT Misc_difference 406 /note= "encoded by GNN"
FT Misc_difference 409 /note= "encoded by GNN"
FT Misc_difference 420 /note= "encoded by NNT"
FT Misc_difference 420 /note= "encoded by GNC"
FT Misc_difference 706 /note= "encoded by NNN"
FT Misc_difference 708 /note= "encoded by CCN"
FT Misc_difference 719 /note= "encoded by CGN"
FT Misc_difference 728 /note= "encoded by CNT"
FT Misc_difference 729 /note= "encoded by GTN"
FT Misc_difference 759.789 /note= "encoded by NNN"
FT Misc_difference 1425 /note= "encoded by GNA"
FT FR2751985-A1.
PN
PD 06-FEB-1998.
PF 01-AUG-1996; 009733.
PR 01-AUG-1996; FR-009733.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI Bach JF, Bousser MG, Joutel A, Tournier Lasserre E;
DR WPI; 98-133137/13.
DR N-PSDB; V57163.
PT Human Notch3 nucleic acids - and methods for identifying
PT pre-disposition to cerebral autosomal dominant arteriopathy with
PT sub-cortical infarcts and leukoencephalopathy
PS Claim 2; Fig 1a-1g; 42pp; French.
CC This sequence represents a partial human notch3 protein, a transmembrane
CC receptor protein involved in lateral inhibition and regulating
CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of the
CC cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy (CADASIL) type. Blocking expression of a mutated
CC Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC protein can be used to treat CADASIL or related disorders.
SQ Sequence 1872 AA;

Query Match 14.1%; Score 82; DB 1; Length 1872;
Best Local Similarity 29.9%; Pred. No. 6.6;
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps 4;

QY 47 IGLEQSVTSRGDLATCPRGFAVTGCTCGSAGSGWDVRAETTC----- 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1234 VGVPCQQ-TPRGPRCACPPG--LSGPFSCRSPGPGSNASCAAPCLHGSGCRPAPLA 1290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 ---HCOACAGMDWTGARC 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1291 PFFRCACA-QGWTGPRC 1306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
R98207
ID R98207 standard; Protein; 60 AA.
AC R98207;
DT 30-DEC-1996 (first entry)
DE Nucleotide used in production of MSH/MoMuLV chimeric sequence.
KW Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
KW polytropic MX27 provirus; targeted drug delivery; gene therapy;
KW single chain antibody; envelope protein; ss.
OS Synthetic.
PN WO9630504-A1.
PD 03-OCT-1996.
```

[illegible]

PS Disclosure: Fig 4B: 37pp; English.
 CC The present sequence is an alternative form of kalinin/laminin 5 gamma-2
 CC chain (see R91427). The gamma-2 chain is of importance to patients
 CC suffering from epidermolysis bullosa, esp. the junctional form (JEB).
 CC Probes and antisense gamma-2 sequences derived from this sequence can be
 CC used to detect, monitor and inhibit the invasive growth of cells in
 CC tissue, partic. malignant tissue.
 SQ Sequence 1111 AA;

Query Match 12.4%; Score 72; DB 1; Length 1111;
 Best Local Similarity 25.4%; Pred. No. 35;
 Matches 31; Conservative 14; Mismatches 47; Indels 30; Gaps 8;
 QY 4 LCLLLPVLGLVSKTLC-----SMEFATNERIQEVAGSLIFRAI-----SSIGLECOVS 54
 Db 10 LCFSLLLPAARATSRREVCDCNCKSRQCIFDRELHROTGN-GFRLNCNDNTDGIHCEK- 68
 QY 55 TSGDLATCPRGF-----AVTGCTCGSAGSWDVRAETTHCHCOC-AGMDWTGARCCRV 106
 Db 68 -----CKNGFYHRRDRCLPCNCKSK-GSLSARCDNSGRCSCKPGV--TGARCDRC 116
 QY 107 QP 108
 Db 117 LP 118

Search completed: October 2, 1999, 11:02:33
 Job time: 197 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 00:40:20 ; Search time 14.11 Seconds
(without alignments)
471.064 Million cell updates/sec

Title: US-09-099-898-2

Perfect score:

Sequence: 1 MKALCLLLPVLGLVSSKT.....CHCQCAGMDWTGARCCRVQP 108

Scoring table: BLOSUM62

Searched: 201082-seqs, 61543640 residues

Database :

SPTREMBL 10:*

```
1: sp archea:*
```

1: sp. archid: *
2: sp bacteria: *

3: sp fungi:*

```
3: sp_range:
4: sp_human: *
```

```
5: sp_invertebrate:*
```

6: sp_mammal:*

```
7: sp_mhc:*
```

```
8: sp_organelle:*
```

9: sp_phage:*

```
10: sp_plant:*
```

```
11: sp_rodent:*
```

```
12: sp_virus:*
```

```
13: sp_vertebrate:*
```

14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	86	14	8	1059	5	P90884	P90884 caenorhabdi
2	84.5	14.5	14.5	1476	13	Q90285	Q90285 carassius a
3	82.5	14.2	14.2	1574	11	O88281	O88281 rattus norv
4	82.5	14.2	14.2	762	13	Q42373	Q42373 brachydanio
5	78	13.4	11.06	5	Q17494	O17494 caenorhabdi	
6	76.5	13.2	589	11	O88671	O88671 rattus norv	
7	76	13	1	2447	13	O13149	O13149 fugu rubrip
8	75.5	13.0	153	4	O75095	O75095 homo sapien	
9	75.5	13.0	387	11	Q06007	Q06007 mus musculu	
10	74.5	12.8	1698	5	Q94438	Q94438 chironomus	
11	74.5	12.8	7.5	5	O96388	O96388 perna virid	
12	74.5	12.8	585	11	O35675	O35675 mus musculu	
13	74.5	12.8	592	11	O88516	O88516 mus musculu	
14	73.5	12.7	830	4	Q43701	Q43701 homo sapien	
15	73.5	12.7	830	4	Q14162	Q14162 homo sapien	
16	73.5	12.7	289	5	P91237	P91237 caenorhabdi	
17	71.5	12.3	378	3	O59925	O59925 penicillium	
18	71.5	12.3	237	4	O75767	O75767 homo sapien	
19	71.5	12.3	105	5	O17187	O17187 bombyx mori	
20	71.5	12.3	1647	5	O45000	O45000 caenorhabdi	
21	71.5	12.3	2704	5	O97458	O97458 drosophila	
22	71	12.2	341	5	Q25032	Q25032 haemochus	
23	71	12.2	752	13	Q42374	Q42374 brachydanio	
24	70.5	12.1	626	10	O49438	O49438 arabidopsis	
25	70	12.0	582	2	O92892	O92892 chlamydia p	
26	70	12.0	709	5	O97444	O97444 giardia lam	
27	70	12.0	1095	11	O60784	O60784 mus musculu	
28	70	12.0	3857	11	O88840	O88840 mus musculu	
29	70	12.0	127	13	O9vgk3	O9vgk3 anquilla ja	

30	69.5	12.0	2796	4	Q95071	Q95071 homo sapien
31	69.5	12.0	1704	5	Q94446	Q94446 chironomus
32	69.5	12.0	3746	6	Q94126	Q94126 ovis aries
33	69.5	12.0	326	10	Q93680	Q93680 persea amer
34	69	11.9	2154	4	Q75443	Q75443 homo sapien
35	69	11.9	266	5	Q94700	Q94700 paramecium
36	69	11.9	625	10	Q65428	Q65428 arabidopsis
37	69	11.9	447	11	Q63348	Q63348 rattus norv
38	69	11.9	379	11	Q35883	Q35883 rattus norv
39	68.5	11.8	258	5	Q44179	Q44179 caenorhabdi
40	68.5	11.8	377	5	Q61699	Q61699 brachioasto
41	68.5	11.8	410	11	Q63720	Q63720 rattus norv
42	68.5	11.8	469	11	Q63721	Q63721 rattus norv
43	68	11.7	388	2	Q53521	Q53521 streptomyce
44	68	11.7	73	5	Q62554	Q62554 mytilus edu
45	68	11.7	619	5	Q01610	Q01610 caenorhabdi

ALIGNMENTS

```

RESULT      1
P90884      PRELIMINARY;          PRT;   1059 AA.
ID           AC
AC           P90884;
DT           01-AUG-1998 (TrEMBLrel. 07, Created)
DC           01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DE           01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DEF         F53B6_2 PROTEIN.
DR           F53B6_2.
OS           Caenorhabditis elegans.
OC           Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
ON           Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis..
[1]
RN           SEQUENCE FROM N.A.
RA           WHITE S.;
RL           Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
[2]
RX           SEQUENCE FROM N.A.
RY           MEDLINE; 94150718.
RZ           WILSON K., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA           BONFELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA           CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA           GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA           JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA           LIGHTNING J., LLOYD C., McMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA           PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA           SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA           THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSON R.,
RA           WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT           "2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT           elegans.";
RD           Nature 368:32-38(1994).
DR           ENBL; Z81086; CAB03121.1; -.
SR           PFAM, PF00047; lg: 1.
SQ           SEQUENCE    1059 AA;  117768 MW;  4CDD1913 CRC32;

Query Match            14.8%; Score 86; DB 5; Length 1059;
Best Local Similarity  27.2%; Pred. No. 0.18;
Matches 31; Conservative 12; Mismatches 33; Indels 38; Gaps

QY       20 TLCSMEEFANE-----RIQEVAGSLIPFRATSSITGLECOSVTSRGDLATCPGFA-VTG 71
        |:|::||::| |::|::| |::|::| |::|::| |::|::| |::|::|
Db        80 TICGGEEIVSRGOCEVFVCRSRTLGTGNFLWRV--DGTFCAAATSr---AVCKSGSCQIVG 134
        -:-|-|-|-|-|-|-|-|-|-|-|WDVRAE-TTCHCCQAG-MDMVTGA 101
        |||||
QY       135 CDGLISSFFFDACG-VCGGGRGDTCNGKFHWKVSEETACASCNCDDVIDWSGA 187
        -----TCGSAGCS-----

RESULT      2
P90285
```

[illegible]

db 357 AGANCERSMNCHELPCYNGGSCDTSRGARTCIQGFPGQCOHRSNEGCSKPCSNGL 416

	Matches	27;	Conservative	9;	Mismatches	31;	Indels	17;	Gaps	5;
QY	36	AGSLIFRAISIGLEQ	-----SVTSRGDIATCPRGFAVTCGT-----	CGS-ACGSWDV	83					
Db	142	AGNCSRMCKELPCYKGGSGCTLTTRGACTCTQGGGPGLCQHRSDGCSKPKCHGGL	201							

QY	84	RAETTC-----HCQCAGMDWTGARC	103	
DB	202	CTEETSYPFHCOCTN-GWKGKRC	224	
RESULT	5			
Q17494				
ID	Q17494	PRELIMINARY;	PRT;	1106 AA.
AC	Q17494;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DE	01-MAY-1999	(TREMBLrel. 10, Last annotation update)		
DT	B0393.5	PROTEIN.		
GN	B0393.5.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;			
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
RC	[1]			
RP	SEQUENCE FROM N.A.			
RA	SULSTON J.;			
RL	Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases			

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans.",
 DR ENBL; 368:32-38(1994).
 DR ENBL; Z37983; CAA86058.1; -.
 DR PFAM; PF00008; EGF; 5.
 DR PROSITE; PS00010; ASX-HYDROXYL; 6.
 DR PROSITE; PS01187; EGF_CA; 5.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 1106 AA; 121137 MW; 2C93CB51 CRC32;
 Query Match 13.4%; Score 78; DB 5; Length 1106;
 Best Local Similarity 31.1%; Pred. No. 1.5;
 Matches 19; Conservative 4; Mismatches 22; Indels 16; Gaps 3;
 QY 59 DLATCPRG-FAVTCCTGSGAGSDVRAETTC-OCAG-----MDWTGAR 102
 Db 772 DLQCPPLRGLPELDCREAGHYGINCESTCHDGSVACDVITGMCPGALCRAGWEGSS 831
 QY 103 C 103
 Db 832 C 832
 RESULT 6
 ID O88671 PRELIMINARY; PRT; 589 AA.
 AC O88671;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE DELTA 3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BOULTER J., GREENFIELD A., WEINMASTER G.;
 RT "Rattus norvegicus mRNA for Delta 3: a putative ligand for Notch.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AF084576; AAC33303.1; -.
 DR PFAM; PF00008; EGF; 5.
 KW Glycoprotein.
 SQ SEQUENCE 589 AA; 61424 MW; 810A3749 CRC32;
 Query Match 13.2%; Score 76.5; DB 11; Length 589.
 Best Local Similarity 28.8%; Pred. No. 1.1;
 Matches 34; Conservative 9; Mismatches 42; Indels 33; Gaps 8;
 QY 21 LC-----SWEAINEIRIOEVAGS--LIFRAISSIGLEQ-----SVTSRGDLATCPRGF-- 68
 Db 249 LCTVPVSTSSCLNSRVSGPAGTGLPGPCDGNPCANGSGCSETPGSEGCACPRGFYG 808
 QY 68 --AVTCTGTC-----GSAC-GSDVRAETTCQCAGMDWTGARC-----CEVQP 208
 Db 309 PRCEVSGVTCADGCFNGGLCVGGEDPDPSAYVCHCPPA---FQGSNCERRVDRCSLP 363
 RESULT 7
 ID O13149 PRELIMINARY; PRT; 2447 AA.
 AC O13149;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE NOTCH 2 (FRAGMENT).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 OC Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.

RN SEQUENCE FROM N.A.
 RP NAKAMURA T., TROMSDALE J.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AB004829; BAA20535.1; -.
 DR PFAM; PF00023; ank; 6.
 DR PFAM; PF00008; EGF; 3.
 DR PROSITE; PS00066; notch; 3.
 DR PROSITE; PS00010; ASX-HYDROXYL; 22.
 DR PROSITE; PS01187; EGF_CA; 20.
 KW Glycoprotein; EGF-like domain.
 ET NON_TER 1
 SQ SEQUENCE 2447 AA; 262542 MW; 3CDA4F7A CRC32;
 Query Match 13.1%; Score 76; DB 13; Length 2447;
 Best Local Similarity 26.7%; Pred. No. 5.9;
 Matches 24; Conservative 2; Mismatches 20; Indels 44; Gaps 6;
 QY 50 EQSVTSRGDLATCPRGFAVTCGTC-----GSACG-----SW 81
 Db 982 ECDSPCKNG-GTCTDGLGTYRCTCPAGYNGQNCQYVNLRCQVRCHNGSCSHGTGATSW 1040
 QY 82 DVRAETTCQCAGMDWTGARC-----CR 105
 Db 1041 -----TCHCT---MGWTGPYCDVPMDSR 1061
 RESULT 8
 ID O75095 PRELIMINARY; PRT; 153 AA.
 AC O75095;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE MEGF6 (FRAGMENT).
 GN MEGF6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 98360089.
 RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
 RT "Identification of high-molecular-weight proteins with multiple
 EGF-like motifs by motif-trap screening.";
 RL Genomics 51:27-34(1998).
 DR ENBL; AB011539; BAA32467.1; -.
 DR PFAM; PF00008; EGF; 2.
 ET NON_TER 1
 SQ SEQUENCE 153 AA; 15579 MW; 1AC64E20 CRC32;
 Query Match 13.0%; Score 75.5; DB 4; Length 153;
 Best Local Similarity 35.7%; Pred. No. 0.34;
 Matches 20; Conservative 1; Mismatches 26; Indels 9; Gaps 3;
 QY 48 GLEQSVTSRGDLATCPRGFAVTCGTCGAGSDVRAETTCQCAGMDWTGARC 103
 Db 51 GAPCDPVTG---LCLCPPG--RSGATCNLCRRGQFGPSCITLHDCGG---GADC 97
 RESULT 9
 ID O06007 PRELIMINARY; PRT; 387 AA.
 AC O06007;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE NOTCH PROTEIN HOMOLOG 1 (NOTCH A PROTEIN) (FRAGMENT).
 GN NOTCH1 OR NOTCH A.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
-RP SEQUENCE FROM N.A.
RC STRAIN=F1 (CBA X C57BL); TISSUE=WHOLE EMBRYO;
RX MEDLINE: 93178563
RA LARDELLI M., LENDAHU U.;
RT "Moth A and moth B--two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
DR EMBL: X68278; CAA48339.1; .
DR MGD; MGI:97363; NOTCH1.
DR PFAM; PF00008; EGF; 6.
DR PFAM; PF00066; notch; 3.
KW Differentiation; Neurogenesis; Repeat.
FT NON_TER 1 387
FT NON_TER 387 387
SQ SEQUENCE 387 AA; 41497 MW; D1FD6C00 CRC32;

Query Match 13.0%; Score 75.5; DB 11; Length 387;
Best Local Similarity 29.1%; Pred. No. 0.93; Mismatches 4; Indels 33; Gaps 5;
Matches 25; Conservative 4; Mismatches 4; Indels 33; Gaps 5;

Qy 48 GLECOV-----TSRGDLATCPRGFAVTGCTC---GSAGSGWDVR 84
Db 141 GRCESVINGCRGKCKNGVCAVASNTARGFCRCPAGF--EGATCENDARTCGSLRCL 198
Qy 85 AETTC-----HCQCAGMDWTGARC 103
Db 199 NGTCSIGSPRPTCLCLG-SFTGPEC 223

RESULT 10
Q94438
ID Q94438 PRELIMINARY; PRT; 1698 AA.
AC Q94438;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE 185 KDA SILK PROTEIN.
GN SPI85.
OS Chironomus pallidivittatus (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Diptera; Nematocera; Chironomoidea; Chironomidae;
OC Chironominae; Chironomus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SALIVARY GLAND;
RA CASE S.T., COX C., BELL W.C., HOFFMAN R.T., MARTIN J., HAMILTON R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U54640; AAA99803.1; .
SQ SEQUENCE 1698 AA; 186164 MW; 85E8E520 CRC32;

Query Match 12.8%; Score 74.5; DB 5; Length 1698;
Best Local Similarity 26.4%; Pred. No. 5.9; Mismatches 25; Indels 33; Gaps 5;
Matches 23; Conservative 6; Mismatches 25; Indels 33; Gaps 5;

Qy 52 QSVTSRGDLATCPRGFAVTGCTC---CGSAC-----GSMDVRAETTCCHCQC--- 94
Db 1355 QTNWQTCOCSPASCTGCTGAQVWCCKKVCYPAQKCKDSPKTDW---ESSCSCQCPKN 1411
Qy 94 -----AGMDWTGARC---CRVOP 108
Db 1412 MRPPGGCNAAGRTWDDATGCSERCAAVP 1438

RESULT 11
Q96388
-ID Q96388 PRELIMINARY; PRT; 75 AA.
AC Q96388;

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE METALLOTHIONEIN.
OS Perna viridis.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytilidae; Perna.
RN [1]
RP SEQUENCE FROM N.A.
RA KHOO H.W., WONG Y.W., PATEL K.H.;
RT "Green mussel (Perna viridis) metallothionein cDNA.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036904; AAD02054.1; .
SQ SEQUENCE 75 AA; 7415 MW; 942586D8 CRC32;

Query Match 12.8%; Score 74.5; DB 5; Length 75;
Best Local Similarity 27.9%; Pred. No. 0.21; Mismatches 17; Conservative 9; Mismatches 18; Indels 17; Gaps 4;
Matches 17; Conservative 9; Mismatches 18; Indels 17; Gaps 4;

Qy 51 COSVTSRGDLATCPRGFAVTGCTCGSAGSGWDVRAETTCCHCAGMDWTGARC---CRVQ 107
Db 5 CNCIETQ--VCIGTGCSEGCRCGDAC-----KCSGGCGCGCGSG-----CKVVKCQ 50
Qy 108 P 108
Db 51 P 51

RESULT 12
O35675
ID O35675 PRELIMINARY; PRT; 585 AA.
AC O35675;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE M-DELTA-Like 3 GENE PRECURSOR.
GN DLL3 OR M-DELTA-Like 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57B16 X DBA; TISSUE=PRIMITIVE STREAK;
RX MEDLINE: 97417575
RA DUNWOODIE S.L., HENRIQUE D.M.P., HARRISON S.M., BEDDINGTON R.S.P.;
RT "Mouse Dll3: a novel divergent Delta gene which may complement the
RT function of other Delta homologues during early pattern formation in
RT the mouse embryo.";
RL Development 124:3065-3076(1997).
DR EMBL; Y11895; CAA72637.1; .
DR MGD; MGI:1096877; DLL3.
DR PFAM; PF00008; EGF; 6.
KW Signal; Glycoprotein.
FT SIGNAL 1 32
FT SIGNAL 32
SQ SEQUENCE 585 AA; 61129 MW; 0F5E6187 CRC32;

Query Match 12.8%; Score 74.5; DB 11; Length 585;
Best Local Similarity 28.8%; Pred. No. 1.9; Mismatches 34; Conservative 8; Mismatches 43; Indels 33; Gaps 8;
Matches 34; Conservative 8; Mismatches 43; Indels 33; Gaps 8;

Qy 21 LC-----SMEAINERQEVN--GSLIFRAISSIGLEQC-----SVTSRGDLATCPRGF-- 68
Db 245 LCTVPVSTSLNSRVPGPASTGCLLPGPDCNGPCANGSGCSETSGSFECACPRGFY 304
Qy 68 ---AVTGCTC-----GSAC--GSMDVRAETTCCHCAGMDWTGARC-----CRVOP 108
Db 305 LRCEVSGVTCADGCPFGNGGLCVGGEDPDSYVCHCP---PGFGSNCCKRVDRCSLAP 359

RESULT 13
O88516
ID O88516 PRELIMINARY; PRT; 592 AA.

Search completed: October 2, 1999, 02:58:58
Job time: 8318 sec

Result No.	Query	Score	Match	Length	DB	ID	Description
1	86	14.8	934	1	C6C_HUMAN	P13671	homo sapien
2	80.5	13.9	456	1	PTFC_BOVIN	P00745	bos taurus
3	75.5	13.0	2139	1	CRB_DROME	P10040	drosophila
4	75.5	13.0	2531	1	NTC1_MOUSE	Q01705	mus musculus
5	74.5	12.8	2444	1	NTC1_HUMAN	P46531	homo sapien
6	74	12.7	124	1	CHR2_BOMMO	P05687	bombyx mori
7	74	12.7	2531	1	NTC1_RAT	Q07008	rattus norv
8	73.5	12.7	660	1	PRIA_BORBU	Q45032	borrelia bu
9	73	12.6	1049	1	ADP1_YEAST	P25371	saccharomyc
10	73	12.6	1700	1	BAR3_CHITE	Q03376	chironomus
11	73	12.6	2437	1	NTC2_BRARE	P46530	brachydanio
12	72	12.4	1193	1	LMG2_HUMAN	Q13753	homo sapien
13	72	12.4	2318	1	NTC3_MOUSE	Q61982	mus musculus
14	71.5	12.3	2703	1	NOTC_DROME	P07207	drosophila
15	71.5	12.3	519	1	TYR2_HUMAN	P40126	homo sapien
16	71.5	12.3	2476	1	ZAN_PIG	Q28983	sus scrofa
17	71	12.2	2907	1	FBN2_MOUSE	Q61555	mus musculus
18	71	12.2	72	1	MT13_MYTED	P80248	mytilus edu
19	71	12.2	397	1	PRLA_LYSEN	P00778	lysobacter
20	71	12.2	1134	1	TIE1_MOUSE	Q06806	mus musculus
21	70	12.0	2871	1	FBN1_MOUSE	Q61554	mus musculus
22	69.5	12.0	2524	1	NOTC_XENLA	P21783	xenopus lae
23	69	11.9	2715	1	G156_PARRP	P13837	paramecium
24	68.5	11.8	581	1	IRF_RAT	Q64716	rattus norv
25	68	11.7	72	1	MT11_MYTED	P80246	mytilus edu
26	68	11.7	458	1	PTTC_RABIT	Q28661	oryctolagus
27	68	11.7	1138	1	TIE1_HUMAN	P35590	homo sapien
28	67	11.5	372	1	AGL_URTDI	P11218	urtica dioi
29	67	11.5	1801	1	LMB2_RAT	P15800	rattus norv
30	67	11.5	72	1	MT12_MYTED	P26215	cochliobolu
31	67	11.5	364	1	PGLR_COCCA	Q06805	bos taurus
32	67	11.5	1136	1	TIE1_BOVIN	Q06805	bos taurus
33	66.5	11.4	2911	1	FBN2_HUMAN	P35556	homo sapien
34	66.5	11.4	182	1	HPRT_RHOCA	P37171	rhodobacter
35	66.5	11.4	71	1	MT21_MYTED	P80251	mytilus edu
36	66.5	11.4	444	1	SHO4_ECOLI	P09748	escherichia
37	66.5	11.4	1416	1	YBN1_CAEEL	Q03610	caenorhabdi
38	66	11.4	177	1	BTC_MOUSE	Q05928	mus musculus
39	66	11.4	2871	1	FBN1_HUMAN	P35555	homo sapien
40	66	11.4	1300	1	IRN_CAVPO	P14617	cavia porce
41	66	11.4	463	1	MFGM_MOUSE	P21956	mus musculus
42	66	11.4	72	1	MT14_MYTED	P80249	mytilus edu
43	66	11.4	309	1	PNAD_PIG	Q28955	sus scrofa

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; J05064; G179704; -
CC EMBL; J05024; G307228; -
CC EMBL; X72177; G825633; -
CC EMBL; J04506; G618466; -
CC PIR; A32109; A32109.
CC PIR; A34235; A34235.
CC PIR; A34372; A34372.
CC MIN; 217050; -
CC PROSITE; PS00279; MAC_PEPFORIN; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; FALSE_NEG.
CC PROSITE; PS01209; LDLRA_1; 1.
CC PROSITE; PS50068; LDLRA_2; 1.
CC PFAM; PF00050; kazal; 1.
CC PFAM; PF00057; ldl_recept_a; 1.
CC PFAM; PF00084; sushi; 2.
CC PFAM; PF00090; tsp_1; 3.
CC HSP; P01130; lafu.
CC COMPLEMENT_PATHWAY; GLYCOPROTEIN; PLASMA; MEMBRANE ATTACK COMPLEX;
KW CYTOLYSIS; SUSHI; REPEAT; SIGNAL; POLYMORPHISM; EGF-LIKE DOMAIN;
KW TRANSMEMBRANE.
FT SIGNAL 1 21
FT CHAIN 22 934
FT REPEAT 22 80
FT REPEAT 81 137
FT DOMAIN 138 175
FT TRANSMEM 331 349
FT TRANSMEM 354 373
FT DOMAIN 517 553
FT REPEAT 562 611
FT DOMAIN 642 934
FT DOMAIN 643 762
FT REPEAT 643 700
FT REPEAT 703 762
FT DOMAIN 766 840
FT DOMAIN 858 934
FT DISULFID 140 151
FT DISULFID 146 164
FT DISULFID 158 173
FT DISULFID 399 420
FT DISULFID 523 539
FT DISULFID 526 541
FT DISULFID 543 552
FT DISULFID 644 686
FT DISULFID 672 699
FT DISULFID 704 746
FT DISULFID 732 761
FT CARBOHYD 324 324
FT CARBOHYD 855 855
FT VARIANT 119 119
SQ SEQUENCE 934 AA; 104843 MW; D03BBD9D CRC32;

Query Match 14.8%; Score 86; DB 1; Length 934;
Best Local Similarity 29.4%; Pred. NO. 0.33;
Matches 30; Conservative 10; Mismatches 36; Indels 26; Gaps 7;

QY 22 CSMEAINERIQVAGSLIFRAISIGLECSQVTSRGDLATCPRGFAVTG---CTCGSAC 78
Db 699 CORTCEIKPVQVEVLITPTPQRLYRIG---ESIE-----LTCPKGFVAGPSRYTCQG-- 749
QY 79 GSW--DVRAETTC-----HQCAGMDWTGARCCRVQP 108
Db 749 NSWTPPEISNLTCEKDTLTKLGHQCLQKQSGSECICMSP 789

RESULT 2
PRTC_BOVIN
ID PRTC_BOVIN STANDARD; PRT; 456 AA.

```

```

AC P00745;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE VITAMIN-K DEPENDENT PROTEIN C PRECURSOR (EC 3.4.21.69)
DE (AUTOPROTHROMBIN IIA) (ANTICOAGULANT PROTEIN C) (FRAGMENT).
GN PROC.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85014826.
RT LONG G.L., BALAGAJE R.M., MCGILLIVRAY R.T.A.;
RT "Cloning and sequencing of liver cDNA coding for bovine protein C.";
RL PROC. NATL. ACAD. SCI. U.S.A. 81:5653-5656(1984).
RN [2]
RP SEQUENCE OF 40-194.
RX MEDLINE; 83007325.
RA FERNLUND P., STENFLO J.;
RT "Amino acid sequence of the light chain of bovine protein C.";
RL J. BIOL. CHEM. 257:12170-12179(1982).
RN [3]
RP REVISION TO 110.
RX MEDLINE; 83169769.
RA DRAKENBERG T., FERNLUND P., ROEPSTORFF P., STENFLO J.;
RT "Beta-hydroxyaspartic acid in vitamin K-dependent protein C.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:1802-1806(1983).
RN [4]
RP SEQUENCE OF 197-456.
RX MEDLINE; 83007326.
RA STENFLO J., FERNLUND P.;
RT "Amino acid sequence of the heavy chain of bovine protein C.";
RL J. BIOL. CHEM. 257:12180-12190(1982).
RN [5]
RP PROCESSING, AND CALCIUM-BINDING DATA.
RX MEDLINE; 83213513.
RA ESMON N.L., DEBAULT L.E., ESMON C.T.;
RT "Proteolytic formation and properties of gamma-carboxyglutamic acid-
domainless protein C.";
RL J. BIOL. CHEM. 258:5548-5553(1983).
RN [6]
RP PROCESSING, AND CALCIUM-BINDING DATA.
RX MEDLINE; 83213514.
RA JOHNSON A.E., ESMON N.L., LAUE T.M., ESMON C.T.;
RT "Structural changes required for activation of protein C are induced
by Ca2+ binding to a high affinity site that does not contain gamma-
carboxyglutamic acid.";
RL J. BIOL. CHEM. 258:5554-5560(1983).
CC -!- FUNCTION: PROTEIN C IS A VITAMIN K-DEPENDENT SERINE PROTEASE THAT
REGULATES BLOOD COAGULATION BY INACTIVATING FACTORS VA AND VIII
IN THE PRESENCE OF CALCIUM IONS AND PHOSPHOLIPIDS.
CC -!- CATALYTIC ACTIVITY: DEGRADATION OF BLOOD COAGULATION FACTORS VA
AND VIII.
CC -!- SUBUNIT: SYNTHESIZED AS A SINGLE CHAIN PRECURSOR, WHICH IS CLEAVED
INTO A LIGHT CHAIN AND A HEAVY CHAIN HELD TOGETHER BY A DISULFIDE
BOND. THE ENZYME IS THEN ACTIVATED BY THROMBIN, WHICH CLEAVES A
TETRAPEPTIDE FROM THE AMINO END OF THE HEAVY CHAIN; THIS
REACTION, WHICH OCCURS AT THE SURFACE OF ENDOTHELIAL CELLS, IS
STRONGLY PROMOTED BY THROMBOMODULIN.
CC -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -!- PTM: THE VITAMIN K-DEPENDENT, ENZYMIC CARBOXYLATION OF SOME
GLU RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND CALCIUM.
CC -!- CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO ANOTHER SITE, BEYOND
THE GLA DOMAIN. THIS GLA-INDEPENDENT BINDING SITE IS NECESSARY FOR
THE RECOGNITION OF THE THROMBIN-THROMBOMODULIN COMPLEX.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

[illegible]

QY 72 -CTC-----GSACGSWDVRAETTCHCQA 94
| | | : : | | |
Db 116 RCDCAEGWEGRFCLHEVRFNSCSAENGCGCAHYCMEEFRRHCSCA 160

```

RESULT 3
CRB_DROME CRB_DROME STANDARD; PRT; 2139 AA.
ID ID
PI0040;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CRUMBS PROTEIN PRECURSOR (95F).
GN CRB.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
[1]
SEQUENCE FROM N.A.
RC STRAIN=OREGON-R; TISSUE=EMBRYO;
RC MEDLINE; 90263104.
RA TEPASS U., THERES C., KNUST E.:
"Crumbs encodes an EGF-like protein expressed on apical membranes of
RT Drosophila epithelial cells and required for organization of
RT epithelia.";
RL CELL 61:787-799(1990).
[2]
SEQUENCE OF 1663-1955 FROM N.A.
RP MEDLINE; 87218337.
RX KNUST E., DIETRICH U., TEPASS U., BREMER K.A., WEIGEL D.,
RA VAESSIN H., CAMPOS-ORTEGA J.A.;
"EGF homologous sequences encoded in the genome of Drosophila
RT melanogaster, and their relation to neurogenic genes.";
RL EMBO J. 6:761-766(1987).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC POLARITY. IT MAY ACT AS A SIGNAL.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M33753; G552087; ALT_SEQ.
DR EMBL; X05144; E1746; -.
DR EMBL; X05144; G929536; -.
DR PIR; B26637; B26637.
DR PIR; A35672; A35672.
DR FLYBASE; FBgn0000368; crb.
DR PROSITE; PS00010; ASX_HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 15.
DR PFAM; PF00008; EGF; 26.
DR PFAM; PF00054; laminin_G; 3.
DR HSP; P00740; l1xa.
DR DR
KW DIFFERENTIATION; REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE;
KW GLYCOPROTEIN; SIGNAL; PHOSPHORYLATION.
SIGN 1 90
FT CHAIN 1 90 CRUMBS PROTEIN.
FT DOMAIN 91 2139 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303 EGF-LIKE 1.
FT DOMAIN 306 343 EGF-LIKE 2.

```

```
FT DOMAIN 348 EGF-LIKE 3. 849 890 BY SIMILARITY.
FT DOMAIN 388 EGF-LIKE 4. 892 901 BY SIMILARITY.
FT DOMAIN 427 EGF-LIKE 5. 908 919 BY SIMILARITY.
FT DOMAIN 464 EGF-LIKE 6. 928 939 BY SIMILARITY.
FT DOMAIN 501 EGF-LIKE 7. 939 957 BY SIMILARITY.
FT DOMAIN 532 EGF-LIKE 8. 957 966 BY SIMILARITY.
FT DOMAIN 545 EGF-LIKE 9. 966 977 BY SIMILARITY.
FT DOMAIN 582 EGF-LIKE 10. 977 995 BY SIMILARITY.
FT DOMAIN 609 EGF-LIKE 11. 995 1009 BY SIMILARITY.
FT DOMAIN 648 EGF-LIKE 12. 1009 1020 BY SIMILARITY.
FT DOMAIN 687 EGF-LIKE 13. 1020 1022 BY SIMILARITY.
FT DOMAIN 725 EGF-LIKE 14. 1022 1231 BY SIMILARITY.
FT DOMAIN 763 EGF-LIKE 15. 1231 1242 BY SIMILARITY.
FT DOMAIN 802 EGF-LIKE 16. 1242 1496 BY SIMILARITY.
FT DOMAIN 840 EGF-LIKE 17. 1496 1505 BY SIMILARITY.
FT DOMAIN 904 EGF-LIKE 18. 1505 1516 BY SIMILARITY.
FT DOMAIN 942 EGF-LIKE 19. 1516 1774 BY SIMILARITY.
FT DOMAIN 980 EGF-LIKE 20. 1774 1783 BY SIMILARITY.
FT DOMAIN 1021 EGF-LIKE 21. 1783 1794 BY SIMILARITY.
FT DOMAIN 1207 EGF-LIKE 22. 1794 1801 BY SIMILARITY.
FT DOMAIN 1481 EGF-LIKE 23. 1801 1812 BY SIMILARITY.
FT DOMAIN 1759 EGF-LIKE 24. 1812 1821 BY SIMILARITY.
FT DOMAIN 1991 EGF-LIKE 25. 1821 1832 BY SIMILARITY.
FT DOMAIN 2030 EGF-LIKE 26. 1832 1850 BY SIMILARITY.
FT DOMAIN 271 EGF-LIKE 27. 1850 1859 BY SIMILARITY.
FT DOMAIN 276 EGF-LIKE 28. 1859 1870 BY SIMILARITY.
FT DOMAIN 291 EGF-LIKE 29. 1870 1889 BY SIMILARITY.
FT DOMAIN 293 BY SIMILARITY. 1889 1903 BY SIMILARITY.
FT DOMAIN 310 BY SIMILARITY. 1903 1914 BY SIMILARITY.
FT DOMAIN 321 BY SIMILARITY. 1914 1930 BY SIMILARITY.
FT DOMAIN 331 BY SIMILARITY. 1930 1939 BY SIMILARITY.
FT DOMAIN 333 BY SIMILARITY. 1939 1941 BY SIMILARITY.
FT DOMAIN 342 BY SIMILARITY. 1941 1950 BY SIMILARITY.
FT DOMAIN 352 BY SIMILARITY. 1950 1968 BY SIMILARITY.
FT DOMAIN 363 BY SIMILARITY. 1968 1977 BY SIMILARITY.
FT DOMAIN 374 BY SIMILARITY. 1977 1988 BY SIMILARITY.
FT DOMAIN 376 BY SIMILARITY. 1988 1995 BY SIMILARITY.
FT DOMAIN 385 BY SIMILARITY. 1995 2008 BY SIMILARITY.
FT DOMAIN 392 BY SIMILARITY. 2008 2017 BY SIMILARITY.
FT DOMAIN 397 BY SIMILARITY. 2017 2028 BY SIMILARITY.
FT DOMAIN 412 BY SIMILARITY. 2028 2037 BY SIMILARITY.
FT DOMAIN 424 BY SIMILARITY. 2037 2046 BY SIMILARITY.
FT DOMAIN 431 BY SIMILARITY. 2046 2055 BY SIMILARITY.
FT DOMAIN 436 BY SIMILARITY. 2055 2064 BY SIMILARITY.
FT DOMAIN 451 BY SIMILARITY. 2064 2073 BY SIMILARITY.
FT DOMAIN 453 BY SIMILARITY. 2073 2082 BY SIMILARITY.
FT DOMAIN 462 BY SIMILARITY. 2082 2091 BY SIMILARITY.
FT DOMAIN 468 BY SIMILARITY. 2091 2100 BY SIMILARITY.
FT DOMAIN 473 BY SIMILARITY. 2100 2109 BY SIMILARITY.
FT DOMAIN 488 BY SIMILARITY. 2109 2118 BY SIMILARITY.
FT DOMAIN 499 BY SIMILARITY. 2118 2127 BY SIMILARITY.
FT DOMAIN 505 BY SIMILARITY. 2127 2136 BY SIMILARITY.
FT DOMAIN 515 BY SIMILARITY. 2136 2145 BY SIMILARITY.
FT DOMAIN 520 BY SIMILARITY. 2145 2154 BY SIMILARITY.
FT DOMAIN 522 BY SIMILARITY. 2154 2163 BY SIMILARITY.
FT DOMAIN 531 BY SIMILARITY. 2163 2172 BY SIMILARITY.
FT DOMAIN 536 BY SIMILARITY. 2172 2181 BY SIMILARITY.
FT DOMAIN 549 BY SIMILARITY. 2181 2190 BY SIMILARITY.
FT DOMAIN 556 BY SIMILARITY. 2190 2200 BY SIMILARITY.
FT DOMAIN 562 BY SIMILARITY. 2200 2209 BY SIMILARITY.
FT DOMAIN 569 BY SIMILARITY. 2209 2218 BY SIMILARITY.
FT DOMAIN 571 BY SIMILARITY. 2218 2227 BY SIMILARITY.
FT DOMAIN 580 BY SIMILARITY. 2227 2236 BY SIMILARITY.
FT DOMAIN 586 BY SIMILARITY. 2236 2245 BY SIMILARITY.
FT DOMAIN 591 BY SIMILARITY. 2245 2254 BY SIMILARITY.
FT DOMAIN 598 BY SIMILARITY. 2254 2263 BY SIMILARITY.
FT DOMAIN 604 BY SIMILARITY. 2263 2272 BY SIMILARITY.
FT DOMAIN 610 BY SIMILARITY. 2272 2281 BY SIMILARITY.
FT DOMAIN 613 BY SIMILARITY. 2281 2290 BY SIMILARITY.
FT DOMAIN 618 BY SIMILARITY. 2290 2300 BY SIMILARITY.
FT DOMAIN 634 BY SIMILARITY. 2300 2309 BY SIMILARITY.
FT DOMAIN 636 BY SIMILARITY. 2309 2318 BY SIMILARITY.
FT DOMAIN 645 BY SIMILARITY. 2318 2327 BY SIMILARITY.
FT DOMAIN 652 BY SIMILARITY. 2327 2336 BY SIMILARITY.
FT DOMAIN 659 BY SIMILARITY. 2336 2345 BY SIMILARITY.
FT DOMAIN 673 BY SIMILARITY. 2345 2354 BY SIMILARITY.
FT DOMAIN 684 BY SIMILARITY. 2354 2363 BY SIMILARITY.
FT DOMAIN 691 BY SIMILARITY. 2363 2372 BY SIMILARITY.
FT DOMAIN 696 BY SIMILARITY. 2372 2381 BY SIMILARITY.
FT DOMAIN 711 BY SIMILARITY. 2381 2390 BY SIMILARITY.
FT DOMAIN 722 BY SIMILARITY. 2390 2400 BY SIMILARITY.
FT DOMAIN 740 BY SIMILARITY. 2400 2409 BY SIMILARITY.
FT DOMAIN 749 BY SIMILARITY. 2409 2418 BY SIMILARITY.
FT DOMAIN 751 BY SIMILARITY. 2418 2427 BY SIMILARITY.
FT DOMAIN 760 BY SIMILARITY. 2427 2436 BY SIMILARITY.
FT DOMAIN 767 BY SIMILARITY. 2436 2445 BY SIMILARITY.
FT DOMAIN 778 BY SIMILARITY. 2445 2454 BY SIMILARITY.
FT DOMAIN 787 BY SIMILARITY. 2454 2463 BY SIMILARITY.
FT DOMAIN 799 BY SIMILARITY. 2463 2472 BY SIMILARITY.
FT DOMAIN 806 BY SIMILARITY. 2472 2481 BY SIMILARITY.
FT DOMAIN 817 BY SIMILARITY. 2481 2490 BY SIMILARITY.
FT DOMAIN 826 BY SIMILARITY. 2490 2500 BY SIMILARITY.
FT DOMAIN 837 BY SIMILARITY. 2500 2509 BY SIMILARITY.
FT DOMAIN 844 BY SIMILARITY. 2509 2518 BY SIMILARITY.
```

Query Match 13.0%; Score 75.5; DB 1; Length 2139;

Best Local Similarity 25.0%; Pred. No. 7.5;

Matches 24; Conservative 10; Mismatches 37; Indels 25; Gaps 4;

QY 16 VSSKTLCSMEEAINEIRIQEVAGSLIFRAISSIGLEQSVTSRGLDLCPCRGFAVTGCTCG 75

DB 346 VNIHPLCQTNPCLNNGACVVG-----GSGALTCE-----CPRGYAGARCEVD 388

QY 76 S-ACGSMDVRAEITC-----HCQCAGMDWTGARC 103

```
Db 389 TDECASQPCQNNGSCIDRINGFSCDCSGTGYTGAF 424

RESULT 4
NTCL_MOUSE STANDARD; PRT; 2531 AA.
AC Q01705;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 13-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).
GN NOTCH1 OR NOTCH.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE: 93194170.
RA FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIADEK P.J., JENKINS N.A.,
RA COPELAND N.G., GRIDLEY T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch.";
RL GENOMICS 15:259-264(1993).
RN [2]
RP SEQUENCE OF 1551-2170 FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE: 93048835.
RA FRANCO DEL AMO F., SMITH D.E., SWIADEK P.J., GENDRON-MAGUIRE M.,
RA GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.";
RL DEVELOPMENT 115:737-744(1992).
RN [3]
RP SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; Z11886; G288503; -.
DR MGD; MG1:97363; NOTCH1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_CA; 21.
DR PFAM; PF00008; EGF; 35.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR HSP; P00740; LIXA.
KW DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 18
FT CHAIN 19 2531
FT DOMAIN 19 1725
FT DOMAIN 1726 1746
FT TRANSMEM 1747 2531
FT DOMAIN 24 1425
FT DOMAIN 1449 1462
FT DOMAIN 1445 1562
FT REPEAT 1445 1480
FT REPEAT 1481 1522
FT REPEAT 1523 1562
FT DOMAIN 1865 2075

Query Match 13.0%; Score 75.5; DB 1; Length 2531;
Best Local Similarity 29.1%; Pred. No. 8.7;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;

QY 48 GLECSQV-----TSRGDIATCPRGFAVTGTC---GSACGSWDVR 84
Db 1301 GRCESVINGCRGCKNGGVCASVNTARGFCRCPAGE--EGATCENDARTCGSLRCL 1358
QY 85 AETTC-----HCOCAGMDWTGARC 103
Db 1359 NGGTCISGPRSPCTCLCLG-SFTGPEC 1383

RESULT 5
NTCL_HUMAN STANDARD; PRT; 2444 AA.
AC P46531;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLLOCATION-
DE ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN NOTCH1 OR TANI.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91347367.
RA ELLISEN L.W., BIRD J., WEST D.C., SORENG A.L., REYNOLDS T.C.,
RA SMITH S.D., SKLAR J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms.";
RL CELL 66:649-661(1991).
CC -1- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; M73980; G338675; -.
DR MIM; 190198; -.
DR PROSITE; PS00010; ASX_HYDROXYL; 20.
DR PROSITE; PS00022; EGF_1; 34.
```

DR .PROSITE: PS01186; EGF_2; 26.
DR PROSITE: PS01187; EGF_CA; 18.
DR PFAM: PF00008; EGF; 35.
DR PFAM: PF00023; ank; 6.
-DR PFAM: PF00066; notch; 3.
KW HSP: P00740; 11XA.
KW DIFFERENTIATION: NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW TRANSMEMBRANE: SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 18
FT CHAIN 19 >2444
FT DOMAIN 19 1736
FT DOMAIN 19 1736
FT TRANSMEM 1737 1757
FT DOMAIN 1758 >2444
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 868
FT DOMAIN 870 906
FT DOMAIN 908 944
FT DOMAIN 946 982
FT DOMAIN 984 1020
FT DOMAIN 1022 1058
FT DOMAIN 1060 1096
FT DOMAIN 1098 1144
FT DOMAIN 1146 1182
FT DOMAIN 1184 1220
FT DOMAIN 1222 1266
FT DOMAIN 1268 1306
FT DOMAIN 1308 1347
FT DOMAIN 1349 1385
FT DOMAIN 1388 1427
FT DOMAIN 1446 1563
FT REPEAT 1446 1481
FT REPEAT 1482 1523
FT REPEAT 1523 1563
FT DOMAIN 1876 2087
FT REPEAT 1876 1921
FT REPEAT 1923 1954
FT REPEAT 1956 1987
FT REPEAT 1990 2021
FT REPEAT 2023 2054
FT REPEAT 2056 2087
FT DOMAIN 1576 1576
FT DOMAIN 1662 1665
FT DOMAIN 1729 1732
FT DOMAIN 1741 1744
FT DOMAIN 1902 1905
FT DOMAIN 2260 2263
FT DOMAIN 2404 2407
FT DOMAIN 2411 2418
FT DOMAIN 24 37
FT DISULFID 31 46
FT DISULFID 48 57
FT DISULFID 63 74
FT DISULFID 68 87
FT DISULFID 89 106
FT DISULFID 106 111
FT DISULFID 111 127
FT DISULFID 127 138
FT DISULFID 138 144
FT DISULFID 144 155
FT DISULFID 155 164
FT DISULFID 164 175
FT DISULFID 175 182
FT DISULFID 182 189
FT DISULFID 189 204
FT DISULFID 204 215
FT DISULFID 215 222
FT DISULFID 222 233
FT DISULFID 233 243
FT DISULFID 243 254
FT DISULFID 254 261
FT DISULFID 261 272
FT DISULFID 272 281
FT DISULFID 281 292
FT DISULFID 292 312
FT DISULFID 312 321
FT DISULFID 321 332
FT DISULFID 332 350
FT DISULFID 350 359
FT DISULFID 359 370
FT DISULFID 370 387
FT DISULFID 387 398
FT DISULFID 398 400
FT DISULFID 400 409
FT DISULFID 409 429
FT DISULFID 429 438
FT DISULFID 438 449
FT DISULFID 449 467
FT DISULFID 467 476
FT DISULFID 476 487
FT DISULFID 487 505
FT DISULFID 505 514
FT DISULFID 514 525
FT DISULFID 525 543
FT DISULFID 543 552
FT DISULFID 552 563
FT DISULFID 563 580
FT DISULFID 580 589
FT DISULFID 589 600
FT DISULFID 600 618
FT DISULFID 618 627
FT DISULFID 627 638
FT DISULFID 638 645
FT DISULFID 645 664
FT DISULFID 664 675
FT DISULFID 675 682
FT DISULFID 682 693
FT DISULFID 693 702
FT DISULFID 702 713
FT DISULFID 713 730
FT DISULFID 730 739
FT DISULFID 739 750
FT DISULFID 750 768
FT DISULFID 768 777
FT DISULFID 777 788
FT DISULFID 788 806
FT DISULFID 806 815
FT DISULFID 815 826
FT DISULFID 826 844
FT DISULFID 844 855
FT DISULFID 855 867
FT DISULFID 867 885
FT DISULFID 885 894
FT DISULFID 894 905
FT DISULFID 905 923
FT DISULFID 923 932
FT DISULFID 932 943
FT DISULFID 943 959
FT DISULFID 959 988
FT DISULFID 988 999
FT DISULFID 999 1008
FT DISULFID 1008 1019
FT DISULFID 1019 1037
FT DISULFID 1037 1046
FT DISULFID 1046 1057
FT DISULFID 1057 1068
FT DISULFID 1068 1081
FT DISULFID 1081 1094
FT DISULFID 1094 1107
FT DISULFID 1107 1120
FT DISULFID 1120 1133
FT DISULFID 1133 1146
FT DISULFID 1146 1159
FT DISULFID 1159 1172
FT DISULFID 1172 1185
FT DISULFID 1185 1198
FT DISULFID 1198 1211
FT DISULFID 1211 1224
FT DISULFID 1224 1237
FT DISULFID 1237 1250
FT DISULFID 1250 1263
FT DISULFID 1263 1276
FT DISULFID 1276 1289
FT DISULFID 1289 1302
FT DISULFID 1302 1315
FT DISULFID 1315 1328
FT DISULFID 1328 1341
FT DISULFID 1341 1354
FT DISULFID 1354 1367
FT DISULFID 1367 1380
FT DISULFID 1380 1393
FT DISULFID 1393 1406
FT DISULFID 1406 1419
FT DISULFID 1419 1432
FT DISULFID 1432 1445
FT DISULFID 1445 1458
FT DISULFID 1458 1471
FT DISULFID 1471 1484
FT DISULFID 1484 1497
FT DISULFID 1497 1510
FT DISULFID 1510 1523
FT DISULFID 1523 1536
FT DISULFID 1536 1549
FT DISULFID 1549 1562
FT DISULFID 1562 1575
FT DISULFID 1575 1588
FT DISULFID 1588 1601
FT DISULFID 1601 1614
FT DISULFID 1614 1627
FT DISULFID 1627 1640
FT DISULFID 1640 1653
FT DISULFID 1653 1666
FT DISULFID 1666 1679
FT DISULFID 1679 1692
FT DISULFID 1692 1705
FT DISULFID 1705 1718
FT DISULFID 1718 1731
FT DISULFID 1731 1744
FT DISULFID 1744 1757
FT DISULFID 1757 1770
FT DISULFID 1770 1783
FT DISULFID 1783 1796
FT DISULFID 1796 1809
FT DISULFID 1809 1822
FT DISULFID 1822 1835
FT DISULFID 1835 1848
FT DISULFID 1848 1861
FT DISULFID 1861 1874
FT DISULFID 1874 1887
FT DISULFID 1887 1900
FT DISULFID 1900 1913
FT DISULFID 1913 1926
FT DISULFID 1926 1939
FT DISULFID 1939 1952
FT DISULFID 1952 1965
FT DISULFID 1965 1978
FT DISULFID 1978 1991
FT DISULFID 1991 2004
FT DISULFID 2004 2017
FT DISULFID 2017 2030
FT DISULFID 2030 2043
FT DISULFID 2043 2056
FT DISULFID 2056 2069
FT DISULFID 2069 2082
FT DISULFID 2082 2095
FT DISULFID 2095 2108
FT DISULFID 2108 2121
FT DISULFID 2121 2134
FT DISULFID 2134 2147
FT DISULFID 2147 2160
FT DISULFID 2160 2173
FT DISULFID 2173 2186
FT DISULFID 2186 2199
FT DISULFID 2199 2212
FT DISULFID 2212 2225
FT DISULFID 2225 2238
FT DISULFID 2238 2251
FT DISULFID 2251 2264
FT DISULFID 2264 2277
FT DISULFID 2277 2290
FT DISULFID 2290 2303
FT DISULFID 2303 2316
FT DISULFID 2316 2329
FT DISULFID 2329 2342
FT DISULFID 2342 2355
FT DISULFID 2355 2368
FT DISULFID 2368 2381
FT DISULFID 2381 2394
FT DISULFID 2394 2407
FT DISULFID 2407 2420
FT DISULFID 2420 2433
FT DISULFID 2433 2446
FT DISULFID 2446 2459
FT DISULFID 2459 2472
FT DISULFID 2472 2485
FT DISULFID 2485 2498
FT DISULFID 2498 2511
FT DISULFID 2511 2524
FT DISULFID 2524 2537
FT DISULFID 2537 2550
FT DISULFID 2550 2563
FT DISULFID 2563 2576
FT DISULFID 2576 2589
FT DISULFID 2589 2602
FT DISULFID 2602 2615
FT DISULFID 2615 2628
FT DISULFID 2628 2641
FT DISULFID 2641 2654
FT DISULFID 2654 2667
FT DISULFID 2667 2680
FT DISULFID 2680 2693
FT DISULFID 2693 2706
FT DISULFID 2706 2719
FT DISULFID 2719 2732
FT DISULFID 2732 2745
FT DISULFID 2745 2758
FT DISULFID 2758 2771
FT DISULFID 2771 2784
FT DISULFID 2784 2797
FT DISULFID 2797 2810
FT DISULFID 2810 2823
FT DISULFID 2823 2836
FT DISULFID 2836 2849
FT DISULFID 2849 2862
FT DISULFID 2862 2875
FT DISULFID 2875 2888
FT DISULFID 2888 2901
FT DISULFID 2901 2914
FT DISULFID 2914 2927
FT DISULFID 2927 2940
FT DISULFID 2940 2953
FT DISULFID 2953 2966
FT DISULFID 2966 2979
FT DISULFID 2979 2992
FT DISULFID 2992 3005
FT DISULFID 3005 3018
FT DISULFID 3018 3031
FT DISULFID 3031 3044
FT DISULFID 3044 3057
FT DISULFID 3057 3070
FT DISULFID 3070 3083
FT DISULFID 3083 3096
FT DISULFID 3096 3109
FT DISULFID 3109 3122
FT DISULFID 3122 3135
FT DISULFID 3135 3148
FT DISULFID 3148 3161
FT DISULFID 3161 3174
FT DISULFID 3174 3187
FT DISULFID 3187 3200
FT DISULFID 3200 3213
FT DISULFID 3213 3226
FT DISULFID 3226 3239
FT DISULFID 3239 3252
FT DISULFID 3252 3265
FT DISULFID 3265 3278
FT DISULFID 3278 3291
FT DISULFID 3291 3304
FT DISULFID 3304 3317
FT DISULFID 3317 3330
FT DISULFID 3330 3343
FT DISULFID 3343 3356
FT DISULFID 3356 3369
FT DISULFID 3369 3382
FT DISULFID 3382 3395
FT DISULFID 3395 3408
FT DISULFID 3408 3421
FT DISULFID 3421 3434
FT DISULFID 3434 3447
FT DISULFID 3447 3460
FT DISULFID 3460 3473
FT DISULFID 3473 3486
FT DISULFID 3486 3499
FT DISULFID 3499 3512
FT DISULFID 3512 3525
FT DISULFID 3525 3538
FT DISULFID 3538 3551
FT DISULFID 3551 3564
FT DISULFID 3564 3577
FT DISULFID 3577 3590
FT DISULFID 3590 3603
FT DISULFID 3603 3616
FT DISULFID 3616 3629
FT DISULFID 3629 3642
FT DISULFID 3642 3655
FT DISULFID 3655 3668
FT DISULFID 3668 3681
FT DISULFID 3681 3694
FT DISULFID 3694 3707
FT DISULFID 3707 3720
FT DISULFID 3720 3733
FT DISULFID 3733 3746
FT DISULFID 3746 3759
FT DISULFID 3759 3772
FT DISULFID 3772 3785
FT DISULFID 3785 3798
FT DISULFID 3798 3811
FT DISULFID 3811 3824
FT DISULFID 3824 3837
FT DISULFID 3837 3850
FT DISULFID 3850 3863
FT DISULFID 3863 3876
FT DISULFID 3876 3889
FT DISULFID 3889 3902
FT DISULFID 3902 3915
FT DISULFID 3915 3928
FT DISULFID 3928 3941
FT DISULFID 3941 3954
FT DISULFID 3954 3967
FT DISULFID 3967 3980
FT DISULFID 3980 3993
FT DISULFID 3993 4006
FT DISULFID 4006 4019
FT DISULFID 4019 4032
FT DISULFID 4032 4045
FT DISULFID 4045 4058
FT DISULFID 4058 4071
FT DISULFID 4071 4084
FT DISULFID 4084 4097
FT DISULFID 4097 4110
FT DISULFID 4110 4123
FT DISULFID 4123 4136
FT DISULFID 4136 4149
FT DISULFID 4149 4162
FT DISULFID 4162 4175
FT DISULFID 4175 4188
FT DISULFID 4188 4201
FT DISULFID 4201 4214
FT DISULFID 4214 4227
FT DISULFID 4227 4240
FT DISULFID 4240 4253
FT DISULFID 4253 4266
FT DISULFID 4266 4279
FT DISULFID 4279 4292
FT DISULFID 4292 4305
FT DISULFID 4305 4318
FT DISULFID 4318 4331
FT DISULFID 4331 4344
FT DISULFID 4344 4357
FT DISULFID 4357 4370
FT DISULFID 4370 4383
FT DISULFID 4383 4396
FT DISULFID 4396 4409
FT DISULFID 4409 4422
FT DISULFID 4422 4435
FT DISULFID 4435 4448
FT DISULFID 4448 4461
FT DISULFID 4461 4474
FT DISULFID 4474 4487
FT DISULFID 4487 4500
FT DISULFID 4500 4513
FT DISULFID 4513 4526
FT DISULFID 4526 4539
FT DISULFID 4539 4552
FT DISULFID 4552 4565
FT DISULFID 4565 4578
FT DISULFID 4578 4591
FT DISULFID 4591 4604
FT DISULFID 4604 4617
FT DISULFID 4617 4630
FT DISULFID 4630 4643
FT DISULFID 4643 4656
FT DISULFID 4656 4669
FT DISULFID 4669 4682
FT DISULFID 4682 4695
FT DISULFID 4695 4708
FT DISULFID 4708 4721
FT DISULFID 4721 4734
FT DISULFID 4734 4747
FT DISULFID 4747 4760
FT DISULFID 4760 4773
FT DISULFID 4773 4786
FT DISULFID 4786 4799
FT DISULFID 4799 4812
FT DISULFID 4812 4825
FT DISULFID 4825 4838
FT DISULFID 4838 4851
FT DISULFID 4851 4864
FT DISULFID 4864 4877
FT DISULFID 4877 4890
FT DISULFID 4890 4903
FT DISULFID 4903 4916
FT DISULFID 4916 4929
FT DISULFID 4929 4942
FT DISULFID 4942 4955
FT DISULFID 4955 4968
FT DISULFID 4968 4981
FT DISULFID 4981 4994
FT DISULFID 4994 5007
FT DISULFID 5007 5020
FT DISULFID 5020 5033
FT DISULFID 5033 5046
FT DISULFID 5046 5059
FT DISULFID 5059 5072
FT DISULFID 5072 5085
FT DISULFID 5085 5098
FT DISULFID 5098 5111
FT DISULFID 5111 5124
FT DISULFID 5124 5137
FT DISULFID 5137 5150
FT DISULFID 5150 5163
FT DISULFID 5163 5176
FT DISULFID 5176 5189
FT DISULFID 5189 5202
FT DISULFID 5202 5215
FT DISULFID 5215 5228
FT DISULFID 5228 5241
FT DISULFID 5241 5254
FT DISULFID 5254 5267
FT DISULFID 5267 5280
FT DISULFID 5280 5293
FT DISULFID 5293 5306
FT DISULFID 5306 5319
FT DISULFID 5319 5332
FT DISULFID 5332 5345
FT DISULFID 5345 5358
FT DISULFID 5358 5371
FT DISULFID 5371 5384
FT DISULFID 5384 5397
FT DISULFID 5397 5410
FT DISULFID 5410 5423
FT DISULFID 5423 5436
FT DISULFID 5436 5449
FT DISULFID 5449 5462
FT DISULFID 5462 5475
FT DISULFID 5475 5488
FT DISULFID 5488 5501
FT DISULFID 5501 5514
FT DISULFID 5514 5527
FT DISULFID 5527 5540
FT DISULFID 5540 5553
FT DISULFID 5553 5566
FT DISULFID 5566 5579
FT DISULFID 5579 5592
FT DISULFID 5592 5605
FT DISULFID 5605 5618
FT DISULFID 5618 5631
FT DISULFID 5631 5644
FT DISULFID 5644 5657
FT DISULFID 5657 5670
FT DISULFID 5670 5683
FT DISULFID 5683 5696
FT DISULFID 5696 5709
FT DISULFID 5709 5722
FT DISULFID 5722 5735
FT DISULFID 5735 5748
FT DISULFID 5748 5761
FT DISULFID 5761 5774
FT DISULFID 5774 5787
FT DISULFID 5787 5800
FT DISULFID 5800 5813
FT DISULFID 5813 5826
FT DISULFID 5826 5839
FT DISULFID 5839 5852
FT DISULFID 5852 5865
FT DISULFID 5865 5878
FT DISULFID 5878 5891
FT DISULFID 5891 5904
FT DISULFID 5904 5917
FT DISULFID 5917 5930
FT DISULFID 5930 5943
FT DISULFID 5943 5956
FT DISULFID 5956 5969
FT DISULFID 5969 5982
FT DISULFID 5982 5995
FT DISULFID 5995 6008
FT DISULFID 6008 6021
FT DISULFID 6021 6034
FT DISULFID 6034 6047
FT DISULFID 6047 6060
FT DISULFID 6060 6073
FT DISULFID 6073 6086
FT DISULFID 6086 6099
FT DISULFID 6099 6112
FT DISULFID 6112 6125
FT DISULFID 6125 6138
FT DISULFID 6138 6151
FT DISULFID 6151 6164
FT DISULFID 6164 6177
FT DISULFID 6177 6190
FT DISULFID 6190 6203
FT DISULFID 6203 6216
FT DISULFID 6216 6229
FT DISULFID 6229 6242
FT DISULFID 6242 6255
FT DISULFID 6255 6268
FT DISULFID 6268 6281
FT DISULFID 6281 6294
FT DISULFID 6294 6307
FT DISULFID 6307 6320
FT DISULFID 6320 6333
FT DISULFID 6333 6346
FT DISULFID 6346 6359
FT DISULFID 6359 6372
FT DISULFID 6372 6385
FT DISULFID 6385 6398
FT DISULFID 6398 6411
FT DISULFID 6411 6424
FT DISULFID 6424 6437
FT DISULFID 6437 6450
FT DISULFID 6450 6463
FT DISULFID 6463 6476
FT DISULFID 6476 6489
FT DISULFID 6489 6502
FT DISULFID 6502 6515
FT DISULFID 6515 6528
FT DISULFID 6528 6541
FT DISULFID 6541 6554
FT DISULFID 6554 6567
FT DISULFID 6567 6580
FT DISULFID 6580 6593
FT DISULFID 6593 6606
FT DISULFID 6606 6619
FT DISULFID 6619 6632
FT DISULFID 6632 6645
FT DISULFID 6645 6658
FT DISULFID 6658 6671
FT DISULFID 6671 6684
FT DISULFID 6684 6697
FT DISULFID 6697 6710
FT DISULFID 6710 6723
FT DISULFID 6723 6736
FT DISULFID 6736 6749
FT DISULFID 6749 6762
FT DISULFID 6762 6775
FT DISULFID 6775 6788
FT DISULFID 6788 6801
FT DISULFID 6801 6814
FT DISULFID 6814 6827
FT DISULFID 6827 6840
FT DISULFID 6840 6853
FT DISULFID 6853 6866
FT DISULFID 6866 6879
FT DISULFID 6879 6892
FT DISULFID 6892 6905
FT DISULFID 6905 6918
FT DISULFID 6918 6931
FT DISULFID 6931 6944
FT DISULFID 6944 6957
FT DISULFID 6957 6970
FT DISULFID 6970 6983
FT DISULFID 6983 6996
FT DISULFID 6996 7009
FT DISULFID 7009 7022
FT DISULFID 7022 7035
FT DISULFID 7035 7048
FT DISULFID 7048 7061
FT DISULFID 7061 7074
FT DISULFID 7074 7087
FT DISULFID 7087 7100
FT DISULFID 7100 7113
FT DISULFID 7113 7126
FT DISULFID 7126 7139
FT DISULFID 7139 7152
FT DISULFID 7152 7165
FT DISULFID 7165 7178
FT DISULFID 7178 7191
FT DISULFID 7191 7204
FT DISULFID 7204 7217
FT DISULFID 7217 7230
FT DISULFID 7230 7243
FT DISULFID 7243 7256
FT DISULFID 7256 7269
FT DISULFID 7269 7282
FT DISULFID 7282 7295
FT DISULFID 7295 7308
FT DISULFID 7308 7321
FT DISULFID 7321 7334
FT DISULFID 7334 7347
FT DISULFID 7347 7360
FT DISULFID 7360 7373
FT DISULFID 7373 7386
FT DISULFID 7386 7399
FT DISULFID 7399 7412
FT DISULFID 7412 7425
FT DISULFID 7425 7438
FT DISULFID 7438 7451
FT DISULFID 7451 7464
FT DISULFID 7464 7477
FT DISULFID 7477 7490
FT DISULFID 7490 7503
FT DISULFID 7503 7516
FT DISULFID 7516 7529
FT DISULFID 7529 7542
FT DISULFID 7542 7555
FT DISULFID 7555 7568
FT DISULFID 7568 7581
FT DISULFID 7581 7594
FT DISULFID 7594 7607
FT DISULFID 7607 7620
FT DISULFID 7620 7633
FT DISULFID 7633 7646
FT DISULFID 7646 7659
FT DISULFID 7659 7672
FT DISULFID 7672 7685
FT DISULFID 7685 7698
FT DISULFID 7698 7711
FT DISULFID 7711 7724
FT DISULFID 7724 7737
FT DISULFID 7737 7750
FT DISULFID 7750 7763
FT DISULFID 7763 7776
FT DISULFID 7776 7789
FT DISULFID 7789 7802
FT DISULFID 7802 7815
FT DISULFID 7815 7828
FT DISULFID 7828 7841
FT DISULFID 7841 7854
FT DISULFID 7854 7867
FT DISULFID 7867 7880
FT DISULFID 7880 7893
FT DISULFID 7893 7906
FT DISULFID 7906 7919
FT DISULFID 7919 7932
FT DISULFID 7932 7945
FT DISULFID 7945 7958
FT DISULFID 7958 7971
FT DISULFID 7971 7984
FT DISULFID 7984 7997
FT DISULFID 7997 8010
FT DISULFID 8010 8023
FT DISULFID 8023 8036
FT DISULFID 8036 8049
FT DISULFID 8049 8062
FT DISULFID 8062 8075
FT DISULFID 8075 8088
FT DISULFID 8088 8101
FT DISULFID 8101 8114
FT DISULFID 8114 8127
FT DISULFID 8127 8140
FT DISULFID 8140 8153
FT DISULFID 8153 8166
FT DISULFID 8166 8179
FT DISULFID 8179 8192
FT DISULFID 8192 8205
FT DISULFID 8205 8218
FT DISULFID 8218 8231
FT DISULFID 8231 8244
FT DISULFID 8244 8257
FT DISULFID 8257 8270
FT DISULFID 8270 8283
FT DISULFID 8283 8296
FT DISULFID 8296 8309
FT DISULFID 8309 8322
FT DISULFID 8322 8335
FT DISULFID 8335 8348
FT DISULFID 8348 8361
FT DISULFID 8361 8374
FT DISULFID 8374 8387
FT DISULFID 8387 8400
FT DISULFID 8400 8413
FT DISULFID 8413 8426
FT DISULFID 8426 8439
FT DISULFID 8439 8452
FT DISULFID 8452 8465
FT DISULFID 8465 8478
FT DISULFID 8478 8491
FT DISULFID 8491 8504
FT DISULFID 8504 8517
FT DISULFID 8517 8530
FT DISULFID 8530 8543
FT DISULFID 8543 8556
FT DISULFID 8556 8569
FT DISULFID 8569 8582
FT DISULFID 8582 8595
FT DISULFID 8595 8608
FT DISULFID 8608 8621
FT DISULFID 8621 8634
FT DISULFID 8634 8647
FT DISULFID 8647 8660
FT DISULFID 8660 8673
FT DISULFID 8673 8686
FT DISUL


```
FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 716 751 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1097 1143 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1307 1346 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1348 1384 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1387 1426 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1449 1462 CYS-RICH.
FT DOMAIN 1464 1499 6 X ANK MOTIF REPEATS.
FT REPEAT 1865 1910 ANK MOTIF 1.
FT REPEAT 1912 1942 ANK MOTIF 2.
FT REPEAT 1944 1975 ANK MOTIF 3.
FT REPEAT 1978 2009 ANK MOTIF 4.
FT REPEAT 2011 2042 ANK MOTIF 5.
FT REPEAT 2044 2076 ANK MOTIF 6.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 286 291 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
```

```
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1112 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.
FT DISULFID 1133 1142 BY SIMILARITY.
FT DISULFID 1149 1160 BY SIMILARITY.
FT DISULFID 1154 1169 BY SIMILARITY.
FT DISULFID 1171 1180 BY SIMILARITY.
FT DISULFID 1187 1198 BY SIMILARITY.
FT DISULFID 1192 1207 BY SIMILARITY.
FT DISULFID 1209 1218 BY SIMILARITY.
FT DISULFID 1225 1244 BY SIMILARITY.
FT DISULFID 1238 1253 BY SIMILARITY.
FT DISULFID 1255 1264 BY SIMILARITY.
FT DISULFID 1271 1284 BY SIMILARITY.
FT DISULFID 1276 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1311 1322 BY SIMILARITY.
FT DISULFID 1316 1334 BY SIMILARITY.
FT DISULFID 1336 1345 BY SIMILARITY.
FT DISULFID 1352 1363 BY SIMILARITY.
FT DISULFID 1357 1372 BY SIMILARITY.
FT DISULFID 1374 1383 BY SIMILARITY.
FT DISULFID 1391 1403 BY SIMILARITY.
```

Query Match 12.7%; Score 74; DB 1; Length 2531;

Best Local Similarity 25.7%; Pred. No. 12;

Matches 36; Conservative 12; Mismatches 52; Indels 40; Gaps 8;

QY 4 LCULLLPVL---GLLVSSKTLCSMEEAINEIRIQVAGSLIFRAISS---IGLECCS----- 54

DB 9 LCILTLPALAARGLSQPS-----GTCINGRCRVEANGTEACVCGAFVQRCODPSPCL 64

QY 54 -----VTSRGDL-----ATCPRGFAVTGC-----TCGSAC-----GSDVRAETT 88

Db 65 STPCNAGTCYVVDHGGIVDYACSPGLGFSPLCLTPLANACLANPCRNCGTCDLLTITE 124
QY 89 CHCOCAGDMWTGARCRCVQP 108
Db 125 YKCRCP-PGWSGKSCQADP 143
RESULT 8
PRIA_BORBU
ID PRIA_BORBU STANDARD; PRT; 660 AA.
AC Q45032; 051047;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
GN PRIA OR B0014.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB19;
RA BOURSAUX-EUDE C.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE; 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi.";
RL NATURE 390:580-586(1997).
CC -1- FUNCTION: PRIA RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX
SSDNA. THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS
PRIB AND PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE
SUBSEQUENT ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN
FUNCTIONS AS A HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X97449; E238797;
DR EMBL; AE001115; G2687882;
DR TIGR; B0014;
DR PRAM; PF00271; helicase-C; 1.
KW DNA REPLICATION; DNA-BINDING; ATP-BINDING; HELICASE; PRIMOSOME;
ZINC-FINGER.
FT NP_BIND 158 165 ATP (POTENTIAL).
FT SITE 256 259 DEEH BOX.
FT ZN_FING 370 382 C4-TYPE (POTENTIAL).
FT ZN_FING 397 413 C4-TYPE (POTENTIAL).
FT CONFLICT 126 126 P -> L (IN REF. 1).
FT CONFLICT 555 555 D -> N (IN REF. 1).
SQ SEQUENCE 660 AA; 77551 MW; B799DC6E CRC32;

Query Match 12.78; Score 73.5; DB 1; Length 660;
Best Local Similarity 29.28; Pred. No. 4.1;
Matches 28; Conservative 13; Mismatches 26; Indels 29; Gaps 7;

QY 16 VSSKTLCSWEENRIQEVAGSLIFRAISSIG-----LECSQSVTSRGDLATCPRGFAVIG 71
Db 337 ISSELSIQLKSNKRO-----SLIF--INKRGYLNKLECNK-----GHICCP-----N 381
QY 72 CTCG-----SAGSGWDVRAETTCCHC-OCAGMD 97
Db 382 CSFGLIYHKKNKLLCHYCSYKTKTASHCPQCESKD 417
RESULT 9
ADP1_YEAST
ID ADP1_YEAST STANDARD; PRT; 1049 AA.
AC P25371;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROBABLE ATP-DEPENDENT PERMEASE PRECURSOR.
GN ADP1 OR YCR011C OR YCR11C OR YCR105.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92160395.
RA PURNELLE B., SKALA J., GOFFEAU A.;
RT "The product of the YCR105 gene located on the chromosome III from
Saccharomyces cerevisiae presents homologies to ATP-dependent
permeases".
RL YEAST 7:867-872(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92327849.
RA SKALA J., PURNELLE B., GOFFEAU A.;
RT "The complete sequence of a 10.8 kb segment distal of SUF2 on the
right arm of chromosome III from Saccharomyces cerevisiae reveals
seven open reading frames including the RVS161, ADP1 and PGK genes".
RL YEAST 8:409-417(1992).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X59720; E264475;
DR PIR; S19421; S19421.
DR PIR; S40914; S40914.
DR SGD; L0000049; ADP1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1.
DR PFAM; PF00008; EGF; 1.
KW ATP-BINDING; TRANSMEMBRANE; GLYCOPROTEIN; TRANSPORT; SIGNAL.
FT SIGNAL 1 25
FT CHAIN 26 1049 PROBABLE ATP-DEPENDENT PERMEASE.
FT NP_BIND 423 430 ATP (BY SIMILARITY).
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 464 481 POTENTIAL.
FT TRANSMEM 794 814 POTENTIAL.
FT TRANSMEM 829 849 POTENTIAL.
FT TRANSMEM 878 898 POTENTIAL.
FT TRANSMEM 910 930 POTENTIAL.
FT TRANSMEM 938 958 POTENTIAL.
FT TRANSMEM 1001 1021 POTENTIAL.
FT TRANSMEM 1025 1045 POTENTIAL.
FT CARBOHYD 50 50 POTENTIAL.
FT CARBOHYD 114 114 POTENTIAL.
FT CARBOHYD 165 165 POTENTIAL.
FT CARBOHYD 221 221 POTENTIAL.

RX MEDLINE: 86079539.
 RA WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene
 RT product that shares homology with proteins containing EGF-like
 RT repeats.";
 RL CELL 43:567-581(1985).
 RN [2].
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE: 87064624.
 RA KIDD S., KELLEY M.R., YOUNG M.W.;
 RT "Sequence of the notch locus of Drosophila melanogaster: relationship
 RT of the encoded protein to mammalian clotting and growth factors.";
 RL MOL. CELL. BIOL. 6:3094-3108(1986).
 RN [3].
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE: 85099329.
 RA WHARTON K.A., YEDVOENICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in D. melanogaster.";
 RL CELL 40:55-62(1985).
 RN [4].
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE: 87257846.
 RA KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
 RT "Restriction of P-element insertions at the Notch locus of Drosophila
 RT melanogaster.";
 RL MOL. CELL. BIOL. 7:1545-1548(1987).
 RN [5].
 RP REVIEW.
 RA HARRIS W.A.;
 RT "Many cell types specified by Notch function.";
 RL CURR. BIOL. 1:120-122(1991).
 CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
 CC ECTODERM.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART
 CC OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS,
 CC THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M16152; G157988; -
 DR EMBL; M16153; G157988; JOINED.
 DR EMBL; M16149; G157988; JOINED.
 DR EMBL; M16150; G157988; JOINED.
 DR EMBL; M16151; G157988; JOINED.
 DR EMBL; K03508; G157993; -
 DR EMBL; M13689; G157993; JOINED.
 DR EMBL; K03507; G157993; JOINED.
 DR EMBL; M12175; G950317; -
 DR EMBL; M16025; G157995; -
 DR PIR; A24420; A24420.
 DR PIR; A24768; A24768.
 DR PIR; A05267; A05267.
 DR FLYBASE; FBgn0004647; N.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 28.
 DR PROSITE; PS01187; EGF_CA; 22.
 DR PFAM; PF00008; EGF; 36.
 DR PFAM; PF00023; ank; 6.

DR PFAM; PF00066; notch; 3.
 DR HSP; P00740; IIXA.
 KW DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
 KW TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
 FT SIGNAL 1 44
 FT CHAIN 45 2703
 FT DOMAIN 45 1745
 FT TRANSMEM 1746 1766
 FT DOMAIN 1767 2703
 FT DOMAIN 58 1451
 FT DOMAIN 58 95
 FT DOMAIN 96 136
 FT DOMAIN 139 176
 FT DOMAIN 177 215
 FT DOMAIN 217 253
 FT DOMAIN 255 291
 FT DOMAIN 293 329
 FT DOMAIN 331 370
 FT DOMAIN 372 408
 FT DOMAIN 409 447
 FT DOMAIN 448 486
 FT DOMAIN 488 524
 FT DOMAIN 526 562
 FT DOMAIN 564 600
 FT DOMAIN 602 637
 FT DOMAIN 639 675
 FT DOMAIN 677 713
 FT DOMAIN 715 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 865
 FT DOMAIN 867 905
 FT DOMAIN 907 944
 FT DOMAIN 946 982
 FT DOMAIN 984 1020
 FT DOMAIN 1022 1058
 FT DOMAIN 1060 1096
 FT DOMAIN 1098 1134
 FT DOMAIN 1136 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1257
 FT DOMAIN 1259 1295
 FT DOMAIN 1297 1335
 FT DOMAIN 1337 1373
 FT DOMAIN 1375 1412
 FT DOMAIN 1415 1451
 FT DOMAIN 1475 1593
 FT REPEAT 1514 1553
 FT REPEAT 1554 1593
 FT REPEAT 1594 2109
 FT DOMAIN 2109 2558
 FT DISULFID 62 73
 FT DISULFID 67 83
 FT DISULFID 85 94
 FT DISULFID 100 111
 FT DISULFID 105 124
 FT DISULFID 126 135
 FT DISULFID 143 154
 FT DISULFID 148 164
 FT DISULFID 166 175
 FT DISULFID 181 192
 FT DISULFID 186 203
 FT DISULFID 205 214
 FT DISULFID 221 232
 FT DISULFID 226 241
 FT DISULFID 243 252
 FT DISULFID 259 270
 FT DISULFID 284 279
 FT DISULFID 281 290
 FT DISULFID 297 308
 FT DISULFID 302 317
 FT DISULFID 319 328

FT CARBOHYD 178 178 POTENTIAL.
 FT CARBOHYD 237 237 POTENTIAL.
 FT CARBOHYD 300 300 POTENTIAL.
 FT CARBOHYD 342 342 POTENTIAL.
 FT CARBOHYD 377 377 POTENTIAL.
 SQ SEQUENCE 519 AA; 59145 MW; 4FEFCDD2 CRC32;

Query Match 12.3%; Score 71.5; DB 1; Length 519;

Best Local Similarity 27.4%; Pred. No. 5.2;

Matches 26; Conservative 7; Mismatches 25; Indels 37; Gaps 4;

Qy 51 COSVTSRGDLATCPRGFAVTGCTCGSAG---SWDVRAET----- 88

Db 29 CMTVDLSVYNKECCPLGAEANVCGSQGRCQCTEVRAETPRWSPGYILRNODDRELWPR 88

Qy 88 -----TCHC--QCAGMD-----WTGARCCRVQP 108

Db 89 KFFHRTCKTGNFAGYNGDCKFGWTGPNCRKKP 123

Search completed: October 2, 1999, 11:06:11
 Job time: 214 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 1999, 10:38:38 ; Search time 12.13 Seconds
(without alignments)
336.725 Million cell updates/sec

Title: US-09-099-898-2
Perfect score: 581
Sequence: 1 MKALCLLLLPVGLLVSSKT.....CHCOCAGMDWTGRCRCRVQP 108

Scoring table: BLOSUM62

Searched: 122810-seqs, 40065486 residues

Database : PIR_60**
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	14.8	934	2 A34372	complement C6 prec
2	82	14.1	2321	2 S78549	notch3 C (activat
3	80.5	13.9	456	1 KXBO	protein C (activat
4	77	13.3	295	2 JC5559	lectin-B - Virgini
5	75.5	13.0	2531	2 A46019	gene Notch-1 prote
6	75.5	13.0	387	2 B49175	Notch A protein -
7	75.5	13.0	2139	2 A35672	Crumbs protein - f
8	75	12.9	124	2 A21761	high-cysteine chor
9	74.5	12.8	2555	2 A40043	notch protein homo
10	74	12.7	2531	2 S18188	notch protein homo
11	73.5	12.7	660	2 F70101	primosomal protein
12	73	12.6	1049	1 S19421	Ang-dependent perm
13	73	12.6	2437	2 S42612	transmembrane prot
14	73	12.6	861	2 A48825	Notch homolog Motc
15	73	12.6	1700	2 S08167	Balbani ring 3 pr
16	72	12.4	2318	2 S45306	notch 3 protein -
17	72	12.4	1111	2 B44018	laminin B2t chain
18	72	12.4	1193	2 A44018	laminin B2t chain
19	71.5	12.3	519	1 YRHU22	dopachrome Delta-i
20	71.5	12.3	2703	2 A24420	notch protein - fr
21	71.5	12.3	105	2 S23061	chordin protein -
22	71	12.2	1134	1 JN0711	protein-tyrosine k
23	71	12.2	396	1 TRYX84	alpha-lytic protei
24	71	12.2	341	2 D48435	cysteine proteinas
25	71	12.2	72	2 S39418	metallothionein 10
26	71	12.2	2907	2 A57278	fibrillin-2 precu
27	70	12.0	2871	2 A55624	fibrillin-1 precu
28	69.5	12.0	346	2 S34165	keratin, type II -
29	69.5	12.0	384	2 S64735	retrovirus-related
30	69.5	12.0	2524	2 A35844	Xotch protein - Af
31	69	11.9	2718	2 A23475	G surface protein
32	69	11.9	447	2 A39321	mucin - rat (fragm
33	68.5	11.8	540	2 B47417	insulin receptor-r
34	68	11.7	1138	1 S24066	protein-tyrosine k
35	68	11.7	388	2 S34672	alkaline serine pr
36	68	11.7	72	2 S39416	metallothionein 10
37	67	11.5	1136	1 S57845	protein-tyrosine k
38	67	11.5	1801	1 MWRTS	laminin beta-2 cha
39	67	11.5	72	2 S39417	metallothionein 10

ALIGNMENTS

RESULT 1
A34372
Complement C6 precursor - human
C:Species: Homo sapiens (man)
C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 07-Aug-1998
C:Accession: A34372; A34235; A31894; A53072
R:Haefliger, J.A.; Tschopp, J.; Vial, N.; Jenne, D.E.
J. Biol. Chem. 264, 18041-18051, 1989
A:Title: Complete primary structure and functional characterization of the sixth comp
A:Reference number: A34372; MUID:90036879
A:Accession: A34372
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-934 <HAE>
A:Cross-references: GB:J05064; NID:g179703; PID:g179704
R:DiScipio, R.G.; Hugli, T.E.
J. Biol. Chem. 264, 16197-16206, 1989
A:Title: The molecular architecture of human complement component C6.
A:Reference number: A34235; MUID:89380223
A:Accession: A34235
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-118, 'E', 120-934 <DIS>
A:Cross-references: GB:J05024; NID:g187824; PID:g307228
R:Chakravarti, D.N.; Chakravarti, B.; Parra, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 86, 2799-2803, 1989
A:Title: Structural homology of complement protein C6 with other channel-forming prot
A:Reference number: A32109; MUID:89202413
A:Accession: A32109
A:Molecule type: mRNA
A:Residues: 1-118, 'E', 120-491 <CH2>
A:Cross-references: GB:J04506; NID:g618465; PID:g618466
R:Chakravarti, D.N.; Muller-Eberhard, H.J.
J. Biol. Chem. 263, 18306-18312, 1988
A:Title: Biochemical characterization of the human complement protein C6. Association
A:Reference number: A31894; MUID:89054009
A:Accession: A31894
A:Molecule type: protein
A:Residues: 22-41 <CHA>
R:Hobart, M.J.; Fernie, B.; DiScipio, R.G.
Biochemistry 32, 6198-6205, 1993
A:Title: Structure of the human C6 gene.
A:Reference number: A53072; MUID:93291175
A:Accession: A53072
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: GGGGGGG, 1-47, 'A', 49-118, 'E', 120-561, 'A', 563-618, 'A', 620-700, 'A', 702-763
A:Note: sequence extracted from NCBI backbone (NCBIP:134071)
C:Genetics:
A:Gene: GDB:C6
A:Cross-references: GDB:119045; OMIM:217050
A:Map position: 5p13-5p13
C:Superfamily: complement c6; agrin inhibitor-like repeat homology; complement factor
gy
C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-934/Product: complement C6 #status experimental <MAT>
F:22-79/Domain: thrombospondin type 1 repeat homology <THRI>
F:80-134/Domain: thrombospondin type 1 repeat homology <THRI>
F:140-173/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:521-552/Domain: EGF homology <EGF>
F:564-612/Domain: thrombospondin type 1 repeat homology <THR3>

F:644-699/Domain: complement factor H repeat homology <PH01>
F:704-761/Domain: complement factor H repeat homology <PH02>

Query Match 14.8%; Score 86; DB 2; Length 934;
Best Local Similarity 29.4%; Pred. No. 0.92;
Matches 30; Conservative 10; Mismatches 36; Indels 26; Gaps 7;

QY 22 CSMEERAINRIQEVAGSLIFRAISSIGLSCQSVTSRGDLATCPRGFAVTG---CTCGSAC 78

Db 699 CQTECIKPKVQEVLTITPQRLYRIG---ESIE-----LTCPKGFVAGPSRYTCQG-- 749

QY 79 GSW--DYRAETTC-----HCCAGMDWTGRCRQVP 108

Db 749 NSWTPPINSITCEKDTLTKLKGHCQ-LGQKQSGSEICMSP 789

RESULT 2

notch3 protein - human

C:Species: Homo sapiens (man)

C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999

C:Accession: S78549; S71825

R:Joutel, A.; Tournier-Lasserre, E.

submitted to the EMBL Data Library, April 1997

A:Reference number: S78549

A:Accession: S78549

A:Molecule type: mRNA

A:Residues: 1-2321 <JOU1>

A:Cross-references: EMBL:U97669; NID:G2668591; PID:G2668592

R:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabrier, H.; Mouton, P.; Alamowit

Nature 383, 707-710, 1996

A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke

A:Reference number: S71825; MUID:97032728

A:Accession: S71825

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2

A:Cross-references: EMBL:U97669

C:Genetics:

A:Gene: notch3

A:Map position: 19p13.1

C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

C:Keywords: tandem repeat, transmembrane protein

F:518-549/Domain: EGF homology <EGF>

F:1838-1870/Domain: ankyrin repeat homology <AN1>

F:1871-1903/Domain: ankyrin repeat homology <AN2>

F:1905-1937/Domain: ankyrin repeat homology <AN3>

F:1938-1970/Domain: ankyrin repeat homology <AN4>

F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 14.1%; Score 82; DB 2; Length 2321;

Best Local Similarity 29.9%; Pred. No. 4.8;

Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps 4;

QY 47 IGLECSVTSRGDLATCPRGFAVTGTCGSACSGWDVRAETTC----- 90

Db 1300 VGPVCQQ-TPRGPRCACPPG--LSGFSRSPGSPPGASNAACAAAPCLHGSGSCRPAPLA 1356

QY 90 ---HCCAGMDWTGARC 103

Db 1357 PFFRCACA-QGWTGPRC 1372

RESULT 3

KXBO

Protein C (activated) (EC 3.4.21.69) precursor - bovine (fragment)

N:Alternate names: autoprothrombin IIA; plasma protein C

C:Species: Bos primigenius taurus (cattle)

QY

Db

QY

Db

C>Date: 30-Nov-1980 #sequence_revision 17-Mar-1987 #text_change 07-Aug-1998
C:Accession: A26250; A18385; A18386; A00928
R:Long, G.L.; Balagaje, R.M.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 81, 5653-5656, 1984
A:Title: Cloning and sequence of liver cDNA coding for bovine protein C.
A:Reference number: A26250; MUID:85014826
A:Accession: A26250
A:Molecule type: mRNA
A:Residues: 1-456 <LON>
R:Fernlund, P.; Stenflo, J.
J. Biol. Chem. 257, 12170-12179, 1982
A:Title: Amino acid sequence of the light chain of bovine protein C.
A:Reference number: A18385; MUID:83007325
A:Accession: A18385
A:Molecule type: protein
A:Residues: 40-194 <FER>
A:Note: 82-Lys was also found
R:Drakenberg, T.; Fernlund, P.; Roepstorff, P.; Stenflo, J.
Proc. Natl. Acad. Sci. U.S.A. 80, 1802-1806, 1983
A:Title: beta-Hydroxyaspartic acid in vitamin K-dependent protein C.
A:Reference number: A19316; MUID:83169769
A:Contents: annotation; revision to residue 110
R:Stenflo, J.; Fernlund, P.
J. Biol. Chem. 257, 12180-12190, 1982
A:Title: Amino acid sequence of the heavy chain of bovine protein C.
A:Reference number: A18386; MUID:83007326
A:Accession: A18386
A:Molecule type: protein
A:Residues: 197-454,'PV' <STE>
R:Esmon, N.L.; Debault, L.E.; Esmon, C.T.
J. Biol. Chem. 258, 5548-5553, 1983
A:Title: Proteolytic formation and properties of gamma-carboxyglutamic acid-domaines
A:Reference number: A37541; MUID:8313513
A:Contents: annotation; activation; calcium binding
R:Johnson, A.E.; Esmon, N.L.; Laue, T.M.; Esmon, C.T.
J. Biol. Chem. 258, 5554-5560, 1983
A:Title: Structural changes required for activation of protein C are induced by Ca2+
A:Reference number: A37542; MUID:83213514
A:Contents: annotation; activation; calcium binding
C:Comment: Protein C is the zymogen of the vitamin K-dependent serine proteinase that
S.
C:Comment: Protein C is synthesized in the liver as a single chain precursor, which i
bin, which cleaves a tetradecapeptide from the amino end of the heavy chain; this rea
C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with s
cognition of the thrombin-thrombomodulin complex.
C:Superfamily: The gamma-carboxyglutamic acid residues arise by a posttranslational, vita
C:Keywords: anticoagulant; beta-hydroxyaspartic acid; blood coagulation; calcium bind
F:1-29/Domain: signal sequence (fragment) #status predicted <SIG>
F:24-83/Domain: Gla domain homology <GLA>
F:30-39/Domain: propeptide #status predicted <PRO>
F:40-194/Product: protein C light chain #status experimental <LCH>
F:98-128/Domain: EGF homology <EGF1>
F:137-172/Domain: EGF homology <EGF2>
F:197-456/Product: protein C heavy chain #status experimental <HCH>
F:197-210/Domain: activation peptide #status experimental <APT>
F:211-440/Domain: trypsin homology <TRY>
F:45/46,53,55,58,59,62,64,65,68,74/Modified site: gamma-carboxyglutamic acid (Glu) #s
F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F:119-128,137-148,144-157,159-172,180-318,237-253,368-382,393-421/Disulfide bonds: #s
F:136,289,350/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:252,298,397/Active site: His, Asp, Ser #status predicted
F:366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9%; Score 80.5; DB 1; Length 456;

Best Local Similarity 24.8%; Pred. No. 1.7;

Matches 26; Conservative 10; Mismatches 40; Indels 29; Gaps 4;

QY 18 SKYLCSMEERAINRIQEVAGSLIFRAISSIGLSCQSVTSRG--DLATCPRGFAVTG---- 72

Db 57 SEEVCFEEA-REIFONTEDTMAVFSYSDQDQCEPRSGSPCDLPCCGKCIDGLGGF 115

```
QY 72 -CTC-----GSACGSWDVRAETTCQCA 94
| | | | | | | | | | | | | | | | | |
Db 116 RDCAGWEGRFLHEVRFNSCNSAENGCGCAHYCMEEGRHCSA 160
| | | | | | | | | | | | | | | | | |

RESULT 4
JC5559
lectin-B - Virginian pokeweed
C:Species: Phytolacca americana (Virginian pokeweed)
C>Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 12-Dec-1997
C:Accession: JC5559
R:Yamaguchi, K.; Yurino, N.; Kino, M.; Ishiguro, M.; Funatsu, G.
BioSci. Biotechnol. Biochem. 61, 690-698, 1997
A:Title: The amino acid sequence of mitogenic lectin-B from the roots of pokeweed (Phytolacca americana)
A:Reference number: JC5559; MUID:97290889
A:Accession: JC5559
A:Molecule type: protein
A:Residues: 1-295 <YAM>
A:Experimental source: root
C:Comment: This protein is a lectin specific for N-acetylglucosamine-containing saccharide
C:Keywords: glycoprotein
F:96,139/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.38; Score 77; DB 2; Length 295;
Best Local Similarity 34.68; Pred. No. 2.4;
Matches 18; Conservative 4; Mismatches 18; Indels 12; Gaps 3;

QY 62 TCPRGFAVTGTCGSACGSWDVRAETTCQCA 104
| | | | | | | | | | | | | | | | | |
Db 223 TCPNELC---CSSGGWGSNDHACGKGCQSDYWRGVDFSGRVCPOGRCC 271
| | | | | | | | | | | | | | | | | |

RESULT 5
A46019
gene Notch-1 protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 14-Aug-1998
C:Accession: A46019
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
Genomics 15, 259-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A:Reference number: A46019; MUID:93194170
A:Accession: A46019
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531 <DEL>
A:Cross-references: GB:211886; GB:S47228; NID:g288502; PID:g288503
A:Note: sequence extracted from NCBI backbone (NCBIP:127318)
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:757-788/Domain: EGF homology <EGF>
F:1917-1948/Domain: ankyrin repeat homology <AN1>
F:1949-1981/Domain: ankyrin repeat homology <AN2>
F:1983-2015/Domain: ankyrin repeat homology <AN3>
F:2016-2048/Domain: ankyrin repeat homology <AN4>
F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 13.08; Score 75.5; DB 2; Length 2531;
Best Local Similarity 29.18; Pred. No. 21;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;

QY 48 GLEQCSV-----TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84
| | | | | | | | | | | | | | | | | |
Db 1301 GRCEVINGCRGPKCKNGGVCASVNTARGFCRCPAGF--EGATCENDARTCGSLRCL 1358
| | | | | | | | | | | | | | | | | |

QY 85 AETTC-----HCQCAGMDWTGARC 103
| | | | | | | | | | | | | | | | | |
Db 1359 NGGTCISGPRSPCTCLCLG-SFTGPEC 1383
| | | | | | | | | | | | | | | | | |

RESULT 6
B49175
```

```
Notch A protein - mouse (fragment)
N:Alternate names: Notch homolog
C:Species: Mus musculus (house mouse)
C>Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 14-Aug-1998
C:Accession: B49175; PH1569; S32109
R:Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 384-372, 1993
A:Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide variety
A:Reference number: A49175; MUID:93178563
A:Accession: B49175
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-387 <LAD>
A:Cross-references: EMBL:X68278; NID:g287987; PID:g287988
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIP:126159)
C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C:Comment: This protein is one of the neurogenic proteins controlling the decision be
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F:27-58/Domain: EGF homology <EGF>

Query Match 13.08; Score 75.5; DB 2; Length 387;
Best Local Similarity 29.18; Pred. No. 4.2;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;

QY 48 GLEQCSV-----TSRGDLATCPRGFAVTGCTC---GSACGSWDVR 84
| | | | | | | | | | | | | | | | | |
Db 141 GRCEVINGCRGPKCKNGGVCASVNTARGFCRCPAGF--EGATCENDARTCGSLRCL 198
| | | | | | | | | | | | | | | | | |

QY 85 AETTC-----HCQCAGMDWTGARC 103
| | | | | | | | | | | | | | | | | |
Db 199 NGGTCISGPRSPCTCLCLG-SFTGPEC 223
| | | | | | | | | | | | | | | | | |

RESULT 7
A35672
crumbs protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 14-Aug-1998
C:Accession: A35672
R:Tepass, U.; Theres, C.; Knust, E.
Cell 61, 787-799, 1990
A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophi
A:Reference number: A35672; MUID:90263104
A:Accession: A35672
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2139 <TEP>
A:Cross-references: GB:M33753
A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for res
C:Genetics:
A:Gene: FlyBase:crb
A:Cross-references: FlyBase:FBgn0000368
C:Superfamily: EGF homology
C:Keywords: transmembrane protein
F:691-722/Domain: EGF homology <EGF>

Query Match 13.08; Score 75.5; DB 2; Length 2139;
Best Local Similarity 25.08; Pred. No. 18;
Matches 24; Conservative 10; Mismatches 37; Indels 25; Gaps 4;

QY 16 VSSKTLCSMEEAINERIOEVAGSLIFRAISSIGLEQCSVTSRGDLATCPRGFAVTGCTCG 75
| | | | | | | | | | | | | | | | | |
Db 346 VNIHPLCOTNPCLNNGACVWIG-----GSGALTC-----CPKGYAGARCEVD 388
| | | | | | | | | | | | | | | | | |

QY 76 S-ACGSWDVRAETTC-----HCQCAGMDWTGARC 103
| | | | | | | | | | | | | | | | | |
Db 389 TDECASQPCQNNNGSCIDRINGFSCDCSGTGTGAF 424
| | | | | | | | | | | | | | | | | |

RESULT 8
```

A;Title: A homolog of Drosophila Notch expressed during mammalian development.
A;Reference number: S18188; MUID:92111383
A;Accession: S18188
A;Molecule type: mRNA
A;Residues: 1-2531 <WEI>
A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology
F;1917-1949/Domain: ankyrin repeat homology <AN1>
F;1950-1982/Domain: ankyrin repeat homology <AN2>
F;1984-2016/Domain: ankyrin repeat homology <AN3>
F;2017-2049/Domain: ankyrin repeat homology <AN4>
F;2050-2082/Domain: ankyrin repeat homology <AN5>

```

Query Match      12.7%; Score 74; DB 2; Length 2531;
Best Local Similarity 25.7%; Pred. No. 29;
Matches 36; Conservative 12; Mismatches 52; Indels 40; Gaps
8

QY 4 LCLLLLPVLT---GLLVSSYTLCSMEAEANERTQEFVAGSLIFRAITS--IGLECOS----- 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 LCLTLLPALAAGRLRCSQPS-----GTCNLNGRCEVANGTACVCSGAFVGRQCDPSPCL 64

QY 54 -----VYTSRDLT-----ATPCRFANVTGC---TCGSAC-----GSWDVRAETT 88
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 STPCKNAGTCVYVDHGGIVDYACSGPLGSPGLCTLPLANACLANPCRNNGGTCDLLTLE 124

```

Db 125 YKCRCP-PGWSKGSCQADP 143

RESULT 11
F70101

C:Species: Borrelia burgdorferi - Lyme disease spirochete
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 18-Sep-1998
C:Accession: F70101
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra,
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.J.; Salzberg, S.; Hanson
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; H
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A::Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A::Reference number: A70100; MUID:98065943 .
A:Accession: F70101
A::Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 660,418

```

Query Match          12.7%; Score 73.5; DB 2; Length 660;
Best Local Similarity 29.2%; Pred. No. 10;
Matches 28; Conservative 13; Mismatches 26; Indels 29; Gaps

QY 16 VSKTLCSMEEANERIQEVAGSLIFRAISSIG-----LEQSVTSRGDIATCPGRFAVYG 71
      :||: | :||::||:| | :||:| | | | | | | | | | | | | | | | | | |
Db 337 ISSLELYSIQKSLNEXRQ---SLIF--INKRGLYNLECNEC---GHIICCP-----N 381

QY 72 CTQG-----SACGSWDVRAETTCCHC-QCAGMD 97

Db 382 CSFGLIYHKKENKILICHVCSYKTKTKTASHCPOCFESKD 417

```

Db 382 CSFGLIYHKENKLLCHYCSYKTKTASHCPQCESKD 417

RESULT 12

S19421

ATP-dependent permease ADP1 precursor - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein YCR011c; protein YCR105
C:Species: *Saccharomyces cerevisiae*

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 17-Jul-1998

C:Accession: S19421; S40914

R:Goffeau, A.; Purnelle, B.; Skala, J.

submitted to the Protein Sequence database, March 1992

A:Reference number: S19420

A:Accession: S19421

A:Molecule type: DNA

A:Residues: 1-1049 <GOF>

A:Cross-references: EMBL:X59720; NID:g1907116; PID:e264475; PID:g1907154; MIPS:YCR011c

R:Purnelle, B.; Skala, J.; Goffeau, A.

Yeast 7, 867-872, 1991

A:Title: The product of the YCR105 gene located on the chromosome III from *Saccharomyces*

A:Reference number: S40914; MUID:92160395

A:Accession: S40914

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1049 <PUR>

R:Skala, J.; Purnelle, B.; Goffeau, A.

Yeast 8, 409-417, 1992

A:Title: The complete sequence of a 10.8 kb segment distal of *SUF2* on the right arm of

K genes.

A:Reference number: S25353; MUID:92327849

A:Contents: annotation

C:Genetics:

A:Gene: SGD:ADP1

A:Cross-references: SGD:S0000604; MIPS:YCR011c

A:Map position: 3R

C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology

C:Keywords: ATP; glycoprotein; P-loop; transmembrane protein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-104/Product: ATP-dependent permease ADP1 #status predicted <MAT>

F:26-324/Domain: extracellular #status predicted <EXT>

F:325-341/Domain: transmembrane #status predicted <TM1>

F:406-607/Domain: ATP-binding cassette homology <ABC>

F:423-430/Region: nucleotide-binding motif A (P-loop)

F:550-557/Region: nucleotide-binding motif B

F:794-810/Domain: transmembrane #status predicted <TM2>

F:878-845/Domain: transmembrane #status predicted <TM3>

F:879-894/Domain: transmembrane #status predicted <TM4>

F:909-925/Domain: transmembrane #status predicted <TM5>

F:938-954/Domain: transmembrane #status predicted <TM6>

F:1025-1041/Domain: transmembrane #status predicted <TM7>

F:50,114,165,221/Binding site: carboxylate (Asn) (covalent) #status predicted

F:429/Binding site: ATP (Lys) #status predicted

Query Match 12.6%; Score 73; DB 1; Length 1049;

Best Local Similarity 35.9%; Pred. No. 17;

Matches 23; Conservative 4; Mismatches 21; Indels 16; Gaps 6;

QY 50 ECQSVTSRGDLATCPRGFAVTGCT---CG-----SAGGSND--VRAET-TCHCOCAGMDWT 99

DB 71 ECNSTGT---CECIEGAGDDCSPLCGGLSPDSGNKDRPIRAQNDTCHCD---NGWG 124

QY 100 GARC 103

DB 125 GINC 128

RESULT 13

S42612

transmembrane protein precursor - zebra fish

C:Species: *Brachydanio rerio* (zebra fish)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 10-Jul-1998

C:Accession: S42612

R:Bierkamp, C.; Campos-Ortega, J.A.

Mech. Dev. 43, 87-100, 1993

A:Title: A zebrafish homologue of the *Drosophila* neurogenic gene Notch and its pattern

A:Reference number: S42612; MUID:94128602

A:Accession: S42612

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2437 <BIE>

A:Cross-references: EMBL:X69088; NID:g433866; PID:g433867

C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology

F:1915-1947/Domain: ankryrin repeat homology <AN1>

F:1948-1980/Domain: ankryrin repeat homology <AN2>

F:1982-2014/Domain: ankryrin repeat homology <AN3>

F:2015-2047/Domain: ankryrin repeat homology <AN4>

F:2048-2080/Domain: ankryrin repeat homology <AN5>

Query Match 12.6%; Score 73; DB 2; Length 2437;

Best Local Similarity 30.6%; Pred. No. 35;

Matches 22; Conservative 3; Mismatches 21; Indels 26; Gaps 5;

QY 51 CQSVTSRGD-----LATCPRGFAVTGCTC-----GSACGSWDVRAETTHC 91

DB 1028 CQNGSCQDGYGYKCTCPHY--TGLNCQSLVRWCDSPPCKNGGSC--WQOGASFTQC 1083

QY 92 QCAGMDWTGARC 103

DB 1084 ---ASGWTGIYC 1092

RESULT 14

A48825

Notch homolog Moch protein - mouse (fragment)

C:Species: *Mus musculus* (house mouse)

C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 14-Aug-1998

C:Accession: A48825

R:Reaume, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Rossant, J.

Dev. Biol. 154, 377-387, 1992

A:Title: Expression analysis of a Notch homologue in the mouse embryo.

A:Reference number: A48825; MUID:93050801

A:Accession: A48825

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-861 <REA>

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:119144)

C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homol

F:26-57/Domain: EGF homology <EGF>

Query Match 12.6%; Score 73; DB 2; Length 861;

Best Local Similarity 30.3%; Pred. No. 14;

Matches 20; Conservative 5; Mismatches 21; Indels 20; Gaps 5;

QY 48 GLEQCQSVTSRGDLATCPRGFAVTGCTGCSACGSWDVRAETTHCQCAGMDWTGARC---- 104

DB 264 GLNCQNLVRWCDSAPCKN-----GGRC--W--QTNTQYHCECRS-GWTGVNCDVLS 309

QY 104 --CRVQ 107

DB 310 VSCEVK 315

RESULT 15

S08167

Balbani ring 3 protein - midge (*Chironomus tentans*)

C:Species: *Chironomus tentans*

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Mar-1998

C:Accession: S08167

R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.

J. Mol. Biol. 211, 331-349, 1990

A:Title: The Balbani ring 3 gene in *Chironomus tentans* has a diverged repetitive str

A:Reference number: S08167; MUID:90172404

A:Accession: S08167

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1700 <PAU>

A:Cross-references: GB:X52263; NID:g7057; PID:g7058

C:Genetics:

A:Gene: BR3

A:Map position: 4

Search completed: October 2, 1999, 10:41:34
Job time: 176 sec